

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

DETEKCE FIBRILACE SÍNÍ ZE SIGNÁLŮ EKG

DETECTION OF ATRIAL FIBRILLATION USING ECG SIGNALS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Vendula Běhunčíková

VEDOUCÍ PRÁCE

SUPERVISOR

doc. Ing. Jiří Kozumplík, CSc.

BRNO 2021

Bakalářská práce

bakalářský studijní program **Biomedicínská technika a bioinformatika**

Ústav biomedicínského inženýrství

Studentka: Vendula Běhunčíková

ID: 211195

Ročník: 3

Akademický rok: 2020/21

NÁZEV TÉMATU:

Detekce fibrilace síní ze signálů EKG

POKYNY PRO VYPRACOVÁNÍ:

1) Prostudujte dostupné metody detekce fibrilace síní (FS) ze signálů EKG a vypracujte jejich stručný přehled. 2) Pro detekci FS realizujte detektor komplexů QRS, popište jeho princip a ověřte jeho spolehlivost. 3) Realizujte varianty detekce FS a vyzkoušejte je na vybraných signálech EKG. 4) Vybranou metodu detekce FS optimalizujte na trénovací databázi, která obsahuje po 50 signálech diagnostikovaných jako sinusový rytmus, fibrilace síní a signály s jinými arytmiemi. Následně svou metodu detekce ověřte na odlišné testovací databázi 150 signálů se stejnou skladbou. Obě databáze jsou dostupné na UBMI. 5) Proveďte statistické vyhodnocení získaných výsledků a vypracujte studii, ve které shrňte a komentujte dosažené výsledky.

DOPORUČENÁ LITERATURA:

[1] KENNEDY A., FINLAY D. D., GULDENRING D., BOND R. R., MORAN K., McLAUGHIN J.: Automated detection of atrial fibrillation using R R intervals and multivariate based classification. Journal of Electrocardiology 49, 2016, pp. 871-876

[2] KRIVOSHEI L., WEBER S., BURKART T., MASELI A., BRASIER N., KUHNE M., CONENE D., HUEBNER T., SEECK A., ECKSTEIN J.: Smart detection of atrial fibrillation. Oxford University Press, European Society of Cardiology, Europace 19, 2017, pp. 753-757

Termín zadání: 8.2.2021

Termín odevzdání: 28.5.2021

Vedoucí práce: doc. Ing. Jiří Kozumplík, CSc.

doc. Ing. Jana Kolářová, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Fibrilace síní je jednou z nejběžnějších poruch srdečního rytmu. V dospělé populaci většina studií uvádí prevalenci fibrilací síní v rozsahu 1-6 %. Prevalence dále roste s věkem. Její včasné odhalení je pro pacienta klíčové. Prevencí je poté možno zabránit závažným stavům. Pro detekci je využíváno mnoho různých metod včetně metod hlubokého učení. Cílem této bakalářské práce je popsat metody automatické detekce fibrilace síní, navrhnout a implementovat vybrané algoritmy a otestovat jejich výsledky na testovacím datasetu. K implementaci byl využit software Matlab R2020b.

KLÍČOVÁ SLOVA

Arytmie, fibrilace síní, detekce QRS, elektrokardiogram, metody detekce fibrilace síní

ABSTRACT

Atrial fibrillation is one of the most common cardiac rhythm disorders. The prevalence of atrial fibrillation is reported at 1-6 % of the adult population. The chances of developing atrial fibrillation increase with age. An early detection of this arrhythmia is a key to prevent more serious conditions. Many ways have been found to detect atrial fibrillation episodes in ECG including deep learning methods. The aim of this bachelor's thesis is to describe the problem of atrial fibrillation and the methods used for detection in the ECG record, design an atrial fibrillation detector and test its results. Detector is implemented using a Matlab R2020b software.

KEYWORDS

Arrhythmia, atrial fibrillation, QRS detection, electrocardiogram, methods of atrial fibrillation detection

BĚHUNČÍKOVÁ, Vendula. *Detekce fibrilace síní ze signálů EKG*. Brno, 2021, 60 s. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství. Vedoucí práce: doc. Ing. Jiří Kozumplík, CSc.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Detekce fibrilace síní ze signálů EKG“ jsem vypracovala samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autorka uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autorky

PODĚKOVÁNÍ

Ráda bych poděkovala vedoucímu práce panu doc. Ing. Jiřímu Kozumplíkovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci. Dále bych ráda poděkovala Ing. Marině Ronzhině, Ph.D. za projevenou ochotu a užitečné připomínky.

Obsah

Úvod	10
1 Fibrilace síní	11
1.1 Charakteristika FIS	11
1.2 EKG obraz	11
2 Metody detekce FIS v signálech EKG	13
2.1 Detekce pomocí vln P	13
2.2 RR intervaly	14
2.2.1 Koeficient rozptylu	14
2.2.2 Střední kvadratická odchylka RR intervalů	15
2.2.3 Normalizovaná absolutní odchylka intervalů RR	15
2.2.4 Mediánová absolutní odchylka	16
2.2.5 Koeficient šikmosti	16
2.2.6 Koeficient špičatosti	17
2.2.7 Shannonova entropie	18
2.2.8 Turning Point Ratio	19
2.2.9 Analýza histogramů a Kolmogorovův-Smirnovův test	19
2.2.10 Poincarého mapa	20
2.3 Metody strojového učení	21
2.3.1 Naivní Bayesův klasifikátor	22
2.3.2 Klasifikační stromy	22
2.3.3 Metoda nejbližšího souseda	23
2.3.4 Metoda podpůrných vektorů	23
2.3.5 Neuronové sítě	24
3 Úvod do praktické části	25
3.1 Použitá data	25
3.2 Parametry pro hodnocení úspěšnosti	25
4 Realizace detektoru QRS	28
4.1 Předzpracování dat	28
4.2 Rozhodovací pravidlo	29
4.3 Výsledky detekce QRS komplexů	29
4.4 Diskuze výsledků detekce QRS komplexů	30

5	Výběr příznaků pro použití vybraných metod	31
5.1	Testy normality	32
5.2	Testy statistické významnosti	33
6	Použití prahovacích metod	36
6.1	Prahy pro jednotlivé příznaky	36
7	Použití SVM klasifikátoru	38
7.1	Vytvoření klasifikátoru	39
7.2	Klasifikace dat	39
8	Použití metody kNN	40
8.1	Vytvoření klasifikátoru	40
8.2	Výběr počtu sousedů	40
8.2.1	Křížová validace	42
8.3	Klasifikace dat	42
9	Dosažené výsledky	43
9.1	Výsledky prahovacích metod	43
9.1.1	Detekce trénovacích dat	43
9.1.2	Detekce testovacích dat	43
9.2	Výsledky SVM klasifikace	44
9.3	Výsledky KNN klasifikace	45
9.4	Vyřazení ostatních arytmií ze souboru dat	47
9.5	Kombinace metod	47
9.6	Diskuze výsledků	48
9.7	Porovnání výsledků s publikovanými studiemi	51
	Závěr	52
	Literatura	54
A	Popis postupu ověření navržených algoritmů	58
A.1	Navržené funkce	58
A.2	Ověření prahovacích metod	58
A.3	Ověření metody SVM	58
A.4	Ověření metody KNN	59
B	Obsah adresáře přiloženého v IS VUT	60

Seznam obrázků

1.1	Ukázka signálu s FIS.	12
2.1	Ukázka signálu s obtížně identifikovatelnými P vlnami.	13
2.2	Ukázka rozdělení dat při různých koeficientech šikmosti.	17
2.3	Ukázka rozdělení dat při různých koeficientech špičatosti.	18
2.4	Ukázka rozdílných Poincarého map dvou signálů.	20
2.5	Struktura neuronové sítě s jednou skrytou vrstvou.	24
4.1	Hrubý odhad kvadrátu obálky filtrovaného signálu.	28
4.2	Signál A00006 s vyznačenými QRS komplexy.	29
4.3	Signál 105 se špatně detekovanými QRS komplexy.	30
5.1	P-graf zobrazující normální rozložení.	32
5.2	P-graf zobrazující nenormální rozložení.	33
5.3	Krabicový graf pro statisticky významný příznak pNN50.	35
5.4	Krabicový graf pro statisticky nevýznamné příznaky.	35
6.1	Křivky parametrů při optimalizaci prahu.	36
6.2	Hodnoty MAD pro tři typy signálů.	37
7.1	Princip vzniku lineárního řešení nelineárního problému [35].	38
8.1	Závislost výsledné klasifikace na počtu sousedů.	41
8.2	Optimalizace počtu sousedů.	41
9.1	Grafy závislostí hodnot statistických parametrů pro jednotlivé křížové validace.	46
9.2	Ukázky správně klasifikovaných signálů.	49
9.3	Ukázky nesprávně klasifikovaných signálů.	50

Seznam tabulek

3.1	Rozložení použitých dat	25
3.2	Obecná matice záměn.	26
4.1	Výsledky testování QRS detektoru.	30
5.1	Výsledky statistických testů.	34
9.1	Výsledky prahovacích metod pro trénovací data.	43
9.2	Matice záměn pro trénovací data pro prahovací metody.	43
9.3	Výsledky prahovacích metod pro testovací data.	44
9.4	Matice záměn pro testovací data pro prahovací metody.	44
9.5	Výsledky metody SVM na testovacích datech.	44
9.6	Matice záměn pro testovací data pro metodu SVM.	45
9.7	Výsledky na testovacích datech pro klasifikátor s $k = 11$	45
9.8	Matice záměn pro testovací data pro metodu KNN.	46
9.9	Výsledky pro redukováná testovací data.	47

Úvod

Fibrilace síní (FIS) je nejčastější srdeční arytmií. Síňové fibrilace se projevují změnou elektrické srdeční aktivity, jsou charakterizovány zcela nepravidelným komorovým rytmem a absencí vln P. Velice typickou je i nepravidelnost RR intervalů. Incidence fibrilací síní v populaci narůstá vzhledem k prodloužení očekávané délky života zvláště u pacientů se srdečním onemocněním. Je důležité znát její různé příčiny, klinické projevy a zvolit vhodnou léčbu farmaky nebo kardioverzí.

Pro pacienta je nejdůležitější včasná detekce FIS. Automatická detekce této arytmie umožňuje včasný zásah, a tím i prevenci dalších obtíží. Nediagnostikovaná a neléčená arytmie může vést k vážným důsledkům, které mohou být pro pacienta fatální.

Cílem této bakalářské práce je prostudovat dostupné metody detekce fibrilace síní v signálech EKG a vypracovat jejich přehled a následně vybrané metody implementovat a otestovat na datech dostupných z UBMI.

Algoritmy pro detekci QRS komplexů a pro vyhodnocování signálů byly programovány v prostředí Matlab R2020b. K zjištění úspěšnosti a dalších parametrů hodnotících klasifikační schopnosti algoritmu byl využit testovací software dostupný z UBMI .

1 Fibrilace síní

Fibrilace síní (dále jen FIS) patří mezi nejčastější poruchu srdečního rytmu, neboli arytmií. Vyskytuje se především u starších pacientů nemocných s kardiovaskulárním onemocněním. V dospělé populaci většina studií uvádí prevalenci FIS v rozsahu 1–6 %, prevalence dále roste s věkem [1, 3].

1.1 Charakteristika FIS

Fibrilace síní je charakterizována zcela nepravidelným komorovým rytmem a absencí vln P. Příčinou může být hypertenze, akutní infarkt myokardu, kardiomyopatie, chlopenní vady a další kardiovaskulární onemocnění, kromě nich pak i například alkohol. Příčina je často idiopatická [1, 2].

Fibrilace můžeme klasifikovat podle doby trvání na [2, 4]:

1. Paroxysmální – fibrilace spontánně skončí a naváže na ni normální sinusový rytmus, obvykle do 48 hodin, maximálně však do 7 dní.
2. Perzistující – jednotlivé epizody jsou delší než 7 dní, popřípadě vyžadují kardioverzi.
3. Permanentní – fibrilace přetrvávají déle než 1 rok a nedaří se navrátit sinusový rytmus.

Léčba je zvolena individuálně na základě etiologie, projevů a rizik spojených s arytmií. Včasné odhalení FIS je velice důležité pro zahájení adekvátní léčby pacienta. Pokud nedojde k včasnému odhalení FIS, může to mít pro pacienta fatální následky. Důležité je především předcházet systémové embolizaci [2].

Při FIS může vznikat stáza krve v oušku levé síně. Tato stáza později vede k tvorbě trombů a následné systémové embolizaci. Největším rizikem je pak cévní mozková příhoda (CMP). Kromě CMP může dojít k embolizaci končetin nebo břišních orgánů. Indikací je podávání warfarinu pacientům s FIS [2].

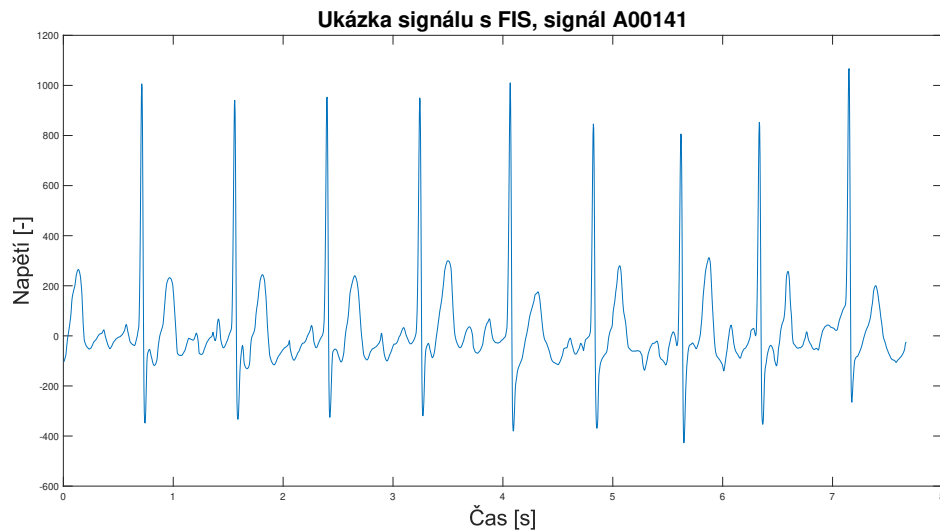
1.2 EKG obraz

Frekvence srdeční aktivity při výskytu FIS vzroste na 350-600/min. Vlnky elektrické aktivity náhodně obíhají srdeční svalovinu a způsobují arytmií. Velmi rychlé a chaotické vzruchy pak znemožňují fyziologické kontrakce síní v každém cyklu, na EKG záznamu pozorujeme nepravidelné vlny (tzv. vlny f) o vysoké frekvenci. Naprosto typická je pak absence vln P v záznamu [2, 3].

Komorový rytmus je pak naprosto nepravidelný a svědčí jednoznačně o FIS nezávisle na rychlosti komorové odpovědi. Vysoká rychlost, až 200/min, se objevuje při fyzické aktivitě, v klidu je častější nízká rychlost. Komorové komplexy QRS jsou

normálně dlouhé, pokud pacient netrpí například raménkovou bloádou či dalšími potížemi. Na EKG záznamu můžeme FIS pozorovat za síňovou extrasystolou či může následovat po flutteru síní [2].

Na obrázku 1.1 je zobrazen úsek signálu A00141 dostupného z CSE databáze obsahující FIS. Signál obsahuje fyziologické QRS komplexy a vlnu T, dále pozorujeme vlny f s různou frekvencí a chybí zde vlny P, nebo jsou těžko rozpoznatelné.



Obr. 1.1: Ukázka signálu s FIS.

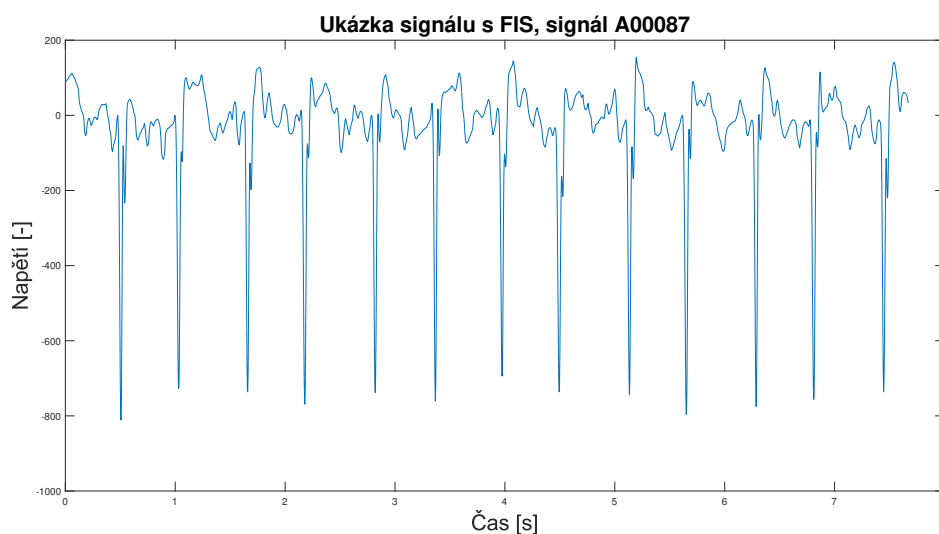
2 Metody detekce FIS v signálech EKG

Pro stanovení léčby je klíčová detekce FIS v EKG záznamu pacienta. Pro detekci můžeme využít různé metody, od méně sofistikovaných po například metody strojového učení, které jsou náročnější pro výběr příznaků i implementaci samotné metody, na druhou stranu mají vyšší účinnost detekce [3].

2.1 Detekce pomocí vln P

Metody detekce FIS pomocí vln P se příliš nevyužívají, jelikož se na EKG záznamu FIS projevují nesynchronními kontrakcemi svalových vláken, v jejichž důsledku P vlna zcela chybí. Navíc se v signálu vyskytují vlny f. Vizuálně se lze přesvědčit, zda se v záznamu nachází či nenachází P vlna, nicméně při větším množství záznamů by bylo vyhodnotit všechny záznamy vizuálně časově velice náročné. Implementace algoritmu hledajícího P vlny je složitá a často nefunkční, nebo funkční s velmi malou úspěšností [3, 4].

Na obrázku 2.1 lze vidět, že navrhovat algoritmus, který by v tomto úseku signálu A00087 dostupného z CSE databáze označoval P vlny, by bylo téměř nemožné.



Obr. 2.1: Ukázka signálu s obtížně identifikovatelnými P vlnami.

P vlna má velice malou amplitudu, nízký poměr šumu a signálu a časové rozdělení P vln často není normální. Různé patologie pak způsobí, že P vlna v signálu buď úplně chybí, je ukrytá v QRS komplexu či ve vlně T. Právě kvůli nízké amplitudě může algoritmus zaměnit vlny f s vlnami P, které by mohly mít stejnou pozici či dokonce amplitudu. V podstatě jakákoli nedostatečná filtrace šumu signálu by pak mohla znemožnit detekci P vlny [3, 4, 7].

U fyziologického záznamu se nachází P vlna před QRS komplexem, čehož se dá využít i při metodách, které vychází z detekce P vln. Je nutné určit časovou oblast, ve které by se měla P vlna nacházet, tedy oddělit z RR intervalu část od konce vlny T po první zákmit QRS komplexu, tedy vlnu Q. V této oblasti by se mělo nacházet maximum, které by znamenalo pozici P vlny. Maximum se dá najít pomocí různých metod, například pomocí adaptivního prahu. Přesněji je možné najít maximum korelací s předpřipravenou šablonou vlny P, vlnkovou transformací či fázorovou transformací [7].

2.2 RR intervaly

Mnohem častější a používanější jsou algoritmy, které fungují na základě analýzy RR intervalů. Pro použití těchto metod je nutné předzpracování signálu v podobě detekce R vln QRS detektorem. RR intervaly jsou intervaly mezi sousedními R vlnami. QRS komplexy jsou v signálech zachovány, ale jsou od sebe různě vzdálené, jinými slovy, intervaly jsou různě dlouhé. Variabilita srdečního rytmu (heart rate variability, dále jen HRV) je často využívaná metrika pro posuzování zdraví pacienta. Krátkodobě může HRV ovlivnit stres organismu, alkohol či léky. Analýza HRV se stala užitečným diagnostickým ukazatelem pro srdeční onemocnění [3, 4].

Tyto metody jsou ve většině případů založené na prahování hodnot daných statisticky významných příznaků. To v praxi znamená, že pokud hodnota příznaku překročí daný práh, je v signálu detekována FIS. Nevýhodou je, že neexistuje žádný univerzální práh, volíme si jej v podstatě libovolně na míru sadě dat, kterou máme k dispozici, což znamená, že pokud by byl stejný algoritmus použit na jinou sadu dat, úspěšnost detekce by mohla poklesnout. Řešením může být například implementovaný adaptivní práh, který by se mohl přizpůsobovat sadě dat, například podle maximální amplitudy, která se v signálu vyskytuje [3, 5].

2.2.1 Koeficient rozptylu

Koeficient rozptylu (dále jen CV) je parametr, jehož hodnota nám může napovědět, zda se v signálu nachází FIS, nebo je signál fyziologický, protože hodnota směrodatné odchylky je vyšší než u signálu s normálním sinusovým rytmem. V případě, že máme dlouhý EKG záznam, signál si rozdělíme do segmentů, ze kterých určíme CV. Lze jej vypočítat pomocí vzorce 2.1

$$CV = \frac{RR_{\sigma}}{RR_{\mu}}, \quad (2.1)$$

kde RR_{σ} je směrodatná odchylka souboru intervalů RR a RR_{μ} je průměr tohoto souboru intervalů [5, 8].

Při výskytu FIS očekáváme vyšší hodnotu CV. Pokud hodnota CV překročí hodnotu prahu, signál je označen za patologický [5, 8].

2.2.2 Střední kvadratická odchylka RR intervalů

Metoda je velice podobná metodě založené na výpočtu CV. Signál se po předchozí detekci R vln rozdělí na segmenty, u nichž pak vypočítáme střední kvadratickou odchylku (dále jen RMSSD) postupných rozdílů RR intervalů pomocí vzorce 2.2

$$RMSSD = \sqrt{\frac{1}{N-1} \left(\sum_{i=1}^{N-1} (RR_{i+1} - RR_i)^2 \right)}, \quad (2.2)$$

kde N je počet RR intervalů v analyzovaném segmentu, RR_i je aktuální RR interval a RR_{i+1} je následující RR interval. Jednotkou RMSSD jsou milisekundy.

Platí, že hodnota RMSSD je vyšší pro signály s FIS než pro signály s normálním sinusovým rytmem. Při analýze pak opět volíme práh, při jehož překročení označíme signál jako patologický [5].

Pro vyhodnocování lze použít i hodnotu takzvané normalizované střední kvadratické odchylky (dále jen nRMSSD). Jedná se o bezrozměrnou hodnotu RMSSD normalizovanou průměrnou hodnotou RR intervalů RR_μ . Přesný vztah pro výpočet nRMSSD uvádí vzorec 2.3

$$nRMSSD = \frac{RMSSD}{RR_\mu}, \quad (2.3)$$

kde RMSSD je střední kvadratická odchylka RR intervalů a RR_μ je průměrná hodnota RR intervalů [5].

2.2.3 Normalizovaná absolutní odchylka intervalů RR

Normalizovaná absolutní odchylka intervalů RR (dále jen NADev) je parametr, který využívá několik intervalů RR. Tento přístup je výpočetně jednoduchý a časově nenáročný, lze díky němu vyvodit nepravidelnost RR intervalů. Signál EKG si rozdělíme na jednotlivé segmenty podle libovolně velkého okna. Ze segmentů musíme vyřadit ektopické cykly, které by mohly zkreslit výsledek analýzy. Interval RR_i vyhovuje pro analýzu, pokud splňuje podmínku 2.4

$$\alpha \cdot RR_\mu < RR_i < \beta \cdot RR_\mu, \quad (2.4)$$

kde α je číslo menší 1, β je číslo větší 1, RR_μ je průměrná hodnota RR intervalů [6].

Pokud ze signálu vybereme vyhovující intervaly RR, můžeme vypočítat hodnotu NADev podle vzorce 2.5

$$NADev = \sum_{i=1}^N \left(\frac{|RR_i - RR_\mu|}{N \cdot RR_\mu} \right), \quad (2.5)$$

kde RR_i je aktuální RR interval, N je počet vyhovujících RR intervalů a RR_μ je průměrná hodnota RR intervalů. Se zvyšující se HRV se hodnota $NADev$ zvyšuje [6].

Vedle parametru $NADev$ existuje parametr normalizované absolutní difference (dále jen $NADiff$). Detekce na základě tohoto příznaku funguje ve své podstatě stejně jako na základě $NADev$ s tím rozdílem, že počítáme s rozdílem dvou po sobě jdoucích RR intervalů, což můžeme vidět ve vzorci 2.6

$$NADiff = \sum_{i=1}^N \left(\frac{|RR_i - RR_{i-1}|}{(N-1) \cdot RR_\mu} \right), \quad (2.6)$$

kde RR_i je aktuální RR interval, RR_{i-1} je předchozí RR interval, RR_μ je průměrná hodnota RR intervalů, N je počet RR intervalů. Platí, že se zvyšující se variabilitou se hodnota $NADiff$ zvyšuje [6].

2.2.4 Mediánová absolutní odchylka

Mediánová absolutní odchylka (dále jen MAD) patří mezi statistický parametr a vyjadřuje variabilitu číselných dat, v tomto případě variabilitu RR intervalů. Platí, že s rostoucí variabilitou dat se zvyšuje hodnota mediánové absolutní odchylky, proto je tento parametr použitelný pro detekci FIS. Signál je nutné rozdělit na segmenty, ve kterých dále probíhá analýza [5].

MAD můžeme vypočítat pomocí vzorce 2.7

$$MAD = median(|RR_i - median(RR_{segment})|), \quad (2.7)$$

kde MAD je mediánová absolutní odchylka RR intervalů, RR_i je jeden RR interval, $median(RR_{segment})$ je medián z konkrétního segmentu signálu [5].

2.2.5 Koeficient šikmosti

Koeficient šikmosti je statistický parametr, jehož hodnota nám podává informace o rozložení dat v datasetu [8, 12].

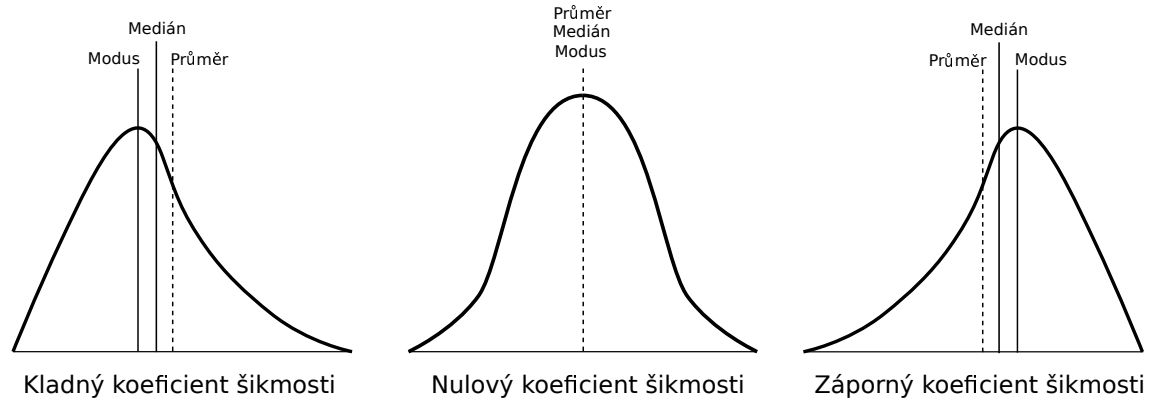
Koeficient šikmosti můžeme vypočítat podle vzorce 2.8

$$koeficient\ šikmosti = \frac{\sum_{i=1}^N (RR_i - RR_\mu)^3}{N \cdot \sigma^3}, \quad (2.8)$$

kde RR_i je RR interval, N je počet RR intervalů a RR_μ je průměrná hodnota RR intervalů, σ je směrodatná odchylka [9].

Nulový koeficient šikmosti značí, že jsou data rozdělena kolem průměru, mediánu i modu symetricky rozloženy, což je graficky znázorněno na obrázku 2.2 vytvořeného podle [12]. Pokud koeficient nabývá hodnoty z intervalu $(-1, -0.5)$, hovoříme o záporné

šikmosti a znamená to, že levá část grafu distribuce je delší a méně strmá, medián bude menší než modus. Při kladných hodnotách koeficientu šikmosti z intervalu (0,5,1) tomu bude přesně naopak. Pravá část grafu distribuce bude delší a méně strmá, medián bude větší než modus [8, 9, 12].



Obr. 2.2: Ukázka rozdělení dat při různých koeficientech šikmosti.

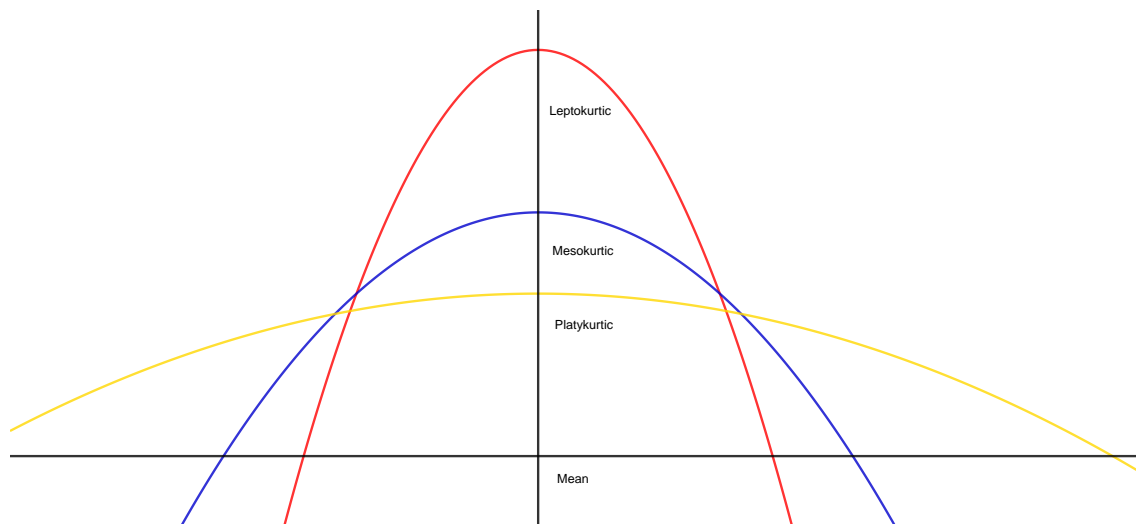
2.2.6 Koeficient špičatosti

Koeficient špičatosti je statistický parametr, který charakterizuje koncentraci hodnot datasetu v blízkosti střední hodnoty. Díky tomu máme představu o rozložení dat [8, 12]. Koeficient špičatosti vypočítáme podle vzorce 2.9

$$koeficient\ spicatosti = \frac{\sum_{i=1}^N (RR_i - RR_\mu)^4}{N \cdot \sigma^4}, \quad (2.9)$$

kde RR_i je RR interval, N je počet RR-intervalů a RR_μ je průměrná hodnota RR intervalů, σ je směrodatná odchylka [9].

Na obrázku 2.3 vytvořeného podle [12] je graficky znázorněno rozdělení dat při různých hodnotách koeficientu špičatosti. Pokud koeficient špičatosti nabývá hodnoty tři, RR intervaly jsou kolem jejich střední hodnoty rovnoměrně rozloženy (Mesokurtic). Při hodnotách vyšších než tři se data koncentrují kolem střední hodnoty, graf distribuce je pak špičatější (Leptokurtic). Naopak, při hodnotách nižších než 3 se data vzdalují střední hodnotě a graf distribuce je naopak plošší (Platykurtic) [12].



Obr. 2.3: Ukázka rozdělení dat při různých koeficientech špičatosti.

2.2.7 Shannonova entropie

Shannonova entropie (dále jen SE), též nazývána jako entropie informační, intuitivně vyjadřuje míru neurčitosti systému. Entropie je průměrné množství informace na jednu hodnotu, množství informace je nezáporné a roste s klesající pravděpodobností výskytu. Pokud nabývá vysokých hodnot, lze říci, že soubor dat je víceméně náhodný a neobsahuje žádné pravidelné vzory. To by se dalo říci například o bílém šumu, nicméně tato metoda se dá aplikovat i na RR-intervaly [10, 11].

Pro výpočet SE je nutné ze signálu vyřadit zvolený počet extrémních hodnot, v tomto případě nazývaných jako ektopické cykly. Tyto cykly by mohly zkreslit výsledky. Posloupnost N intervalů se následně překvantuje na q úrovní, výsledná posloupnost pak nabývá hodnot od jedné do počtu úrovní [10, 13].

SE pak vypočítáme podle vzorce 2.10

$$SE = - \sum_{i=1}^q p(x_i) \cdot \log_z(p(x_i)), \quad (2.10)$$

kde SE je Shannonova entropie, x_i je hodnota náležící do jedné skupiny hodnot, q je počet kvantovacích úrovní, $p(x_i)$ je pravděpodobnost výskytu hodnoty z daného intervalu hodnot [10].

Jak již bylo zmíněno, hodnota SE je pro signály, které jsou více nepravidelné v důsledku nepravidelného vedení vzruchů ze síní do komor a neobsahují žádné opakující se vzory vyšší, než u signálů s normálním sinusovým rytmem [10, 11, 13].

2.2.8 Turning Point Ratio

Poměr kritických bodů (anglicky Turning Point Ratio, dále jen TPR) vyjadřuje počet kritických bodů v daném časovém úseku signálu [10].

Každý RR interval je porovnáván se svými dvěma sousedními RR intervaly. Jako bod zlomu (či kritický bod) je označen, pokud splňuje podmínku 2.11 nebo 2.12

$$RR_{i-1} > RR_i < RR_{i+1}, \quad (2.11)$$

$$RR_{i-1} < RR_i > RR_{i+1}, \quad (2.12)$$

tedy pokud je konkrétní RR-interval RR_i menší než jeho sousední RR-intervaly, nebo naopak větší než jeho dva sousední RR-intervaly [10, 16].

Samotný poměr TPR se dále vypočítá podle vzorce 2.13

$$TPR = \frac{N_{TP}}{N}, \quad (2.13)$$

kde N_{TP} je celkový počet kritických bodů v segmentu a N označuje počet segmentů záznamu [16].

Pro patologické signály obsahující FIS je typická vyšší hodnota TPR než pro signály s normálním sinusovým rytmem. Signály lze vyhodnocovat na základě překročení předem zvoleného prahu. Vysoká hodnota TPR stanovuje míru náhodnosti série RR-intervalů [10, 16].

2.2.9 Analýza histogramů a Kolmogorovův-Smirnovův test

Hustota histogramu je ovlivněna pravidelností RR intervalů. U normálního sinusového rytmu je počet nenulových hodnot hustoty histogramů nižší, obvykle se nachází jeden vysoký peak, zatímco u histogramu signálu s fibrilacemi je rozložení hustoty do většího množství menších peaků. Pokud výška překročí zvolený vysoký práh, je velice pravděpodobné, že se jedná o fyziologický signál, který neobsahuje nepravidelné RR intervaly [8, 17].

Kolmogorovův-Smirnovův statistický test slouží k získání p-hodnoty, která značí podobnost dvou histogramů, konkrétně histogramu analyzovaného signálu a šablony standardního histogramu pro FIS či normální sinusový rytmus. Nulová hypotéza říká, že dva výběry dat mají stejná rozdělení [8, 17].

P-hodnota se Kolmogorovova-Smirnovova testu se vypočítá podle vzorce 2.14

$$p = 2 \cdot \sum_{i=1}^{\infty} (-1)^{i-1} \cdot e^{-2 \cdot i^2 \cdot \lambda^2}, \quad (2.14)$$

pro λ platí vztah 2.15

$$\lambda = \left(\frac{\sqrt{N_e} + 0.12 + 0.11}{\sqrt{N_e}} \right) \cdot D, \quad (2.15)$$

kde D je největší vzdálenost kumulativních četností v grafu obou rozdělení a pro N_e platí vztah 2.16

$$N_e = \frac{N_1 \cdot N_2}{N_1 + N_2}, \quad (2.16)$$

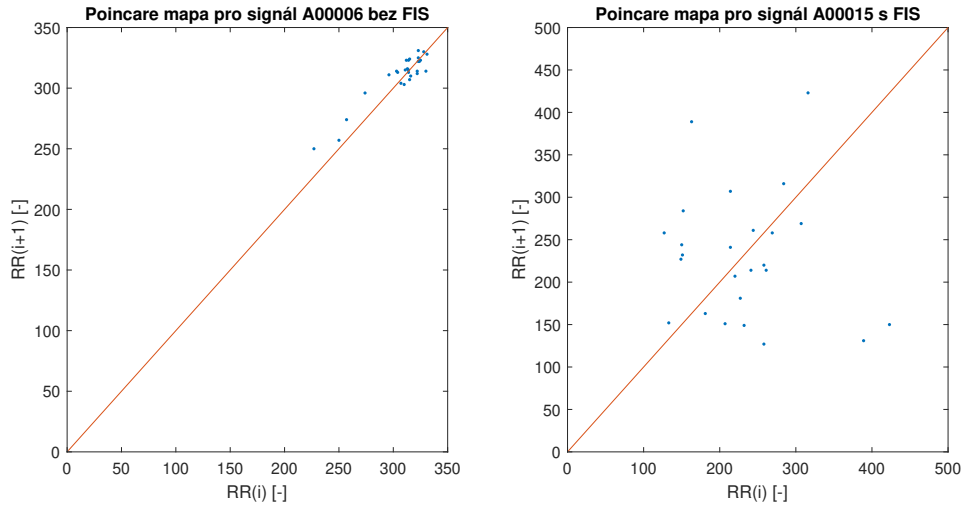
kde N_1 a N_2 jsou celkové počty prvků výběrů [8, 17].

Nízká p-hodnota zamítá nulovou hypotézu o shodném rozložení dat a odpovídá velkému rozdílu mezi histogramy, vysoká hodnota značí podobnost distribuce dat [8, 17].

2.2.10 Poincarého mapa

Poincarého mapa (známá také jako Lorenzova mapa) je graf závislosti RR intervalu na předchozím RR intervalu, znázorňuje časovou změnu RR intervalu. Jednotlivé dvojice jsou ve výsledném grafu zobrazeny jako body [14, 15].

V tomto grafu je následně analyzováno rozložení bodů podél hlavní diagonály. Zajímavé jsou také jednotlivé shluky, které na mapě vznikly. U zdravého člověka se zpravidla nachází pouze jeden shluk a body se nachází v okolí hlavní diagonální linie. Více shluků by pak mohlo indikovat FIS v signálu. Rozdíl mezi Poincarého mapou signálu A00006 z trénovacího datasetu bez FIS a signálu A00015 taktéž z trénovacího datasetu s FIS je zobrazen na obrázku 2.4 [14, 15].



Obr. 2.4: Ukázka rozdílných Poincarého map dvou signálů.

Kromě počtu shluků jsou důležitými parametry hodnoty krátkodobé variability RR intervalů $SD1$ a dlouhodobé variability RR intervalů $SD2$, které lze vypočítat ze vzorců 2.17, 2.18

$$SD1 = \frac{\sqrt{2}}{2} \cdot SD(RR_i - RR_{i+1}), \quad (2.17)$$

$$SD2 = \sqrt{2SD(RR_i)^2 - \frac{1}{2}SD(RR_i - RR_{i+1})^2}, \quad (2.18)$$

kde SD1 a SD2 jsou variability RR-intervalů, RR_i je RR-interval, RR_{i+1} je následující interval, SD je směrodatná odchylka časového záznamu [14].

Poměr SD1/SD2 je při FIS vysoký v důsledku zvýšené variability RR-intervalů, proto je možné rozhodovat na základě překročení zvoleného prahu [11, 14, 15].

2.3 Metody strojového učení

Pro zpřesnění automatické detekce FIS se nyní stále více využívá tzv. metod strojového učení. Tyto metody umožňují rychlejší a přesnější automatické vyhodnocování i velkého množství dat [18].

Obecně můžeme tyto metody v praxi rozdělit podle přístupu k učení na [18]:

- metody s učením bez učitele,
- metody s učením s učitelem.

Metody s učením bez učitele jsou obecně složitější pro implementaci, protože stroji nejsou dodány žádné apriorní znalosti z trénovací množiny, vychází tak pouze z vnitřních souvislostí v datech. Pro využití těchto metod je nutné nejdříve v parametrickém prostoru odhalit třídy objektů, například pomocí shlukovací analýzy [18, 19].

Metody s učením s učitelem stroji předkládají objekty se známou výslednou klasifikací, které si stroj zapamatuje a na základě této zkušenosti je schopen dále zařazovat nové objekty do daných tříd. V případě detekce FIS by byla stroji poskytnuta trénovací množina se známými referencemi k signálům, na základě těchto referencí by pak stroj označoval další signály s FIS. Metody s učitelem jsou nicméně poměrně časově náročné a hrozí zde riziko, že učitel zařadí objekt do špatné skupiny [18, 19].

Metody strojového učení řeší klasifikační problém. Klasifikační algoritmy jsou využívány pro rozdělení určitého datasetu do konečného počtu klasifikačních skupin, přičemž všechny objekty v jednotlivé podmnožině mají dostatečně podobné společné vlastnosti, ideálně úplně totožné. V případě detekce FIS podmnožinami lze rozumět konkrétní kategorie, například klasifikační skupinu označující FIS, skupinu označující fyziologické signály s normálním sinusovým rytmem, případně i skupinu, do které spadají signály s ostatními arytmiemi [18, 19].

Klasifikaci provádí klasifikátor, tedy konkrétní algoritmus, do něž vstupují jednotlivé příznaky, na základě kterých dochází k rozdělení dat do skupin [18, 19].

Základem pro tyto metody je vhodný výběr příznaků, které můžeme vybírat například na základě výsledků statistických testů významnosti daného příznaku,

užitečnou se jeví i korelační matice, která může pomoci vyřadit příliš korelované příznaky [18, 19].

2.3.1 Naivní Bayesův klasifikátor

Naivní Bayesův klasifikátor vychází z Bayesovy věty, pro klasifikaci jsou využívány pravděpodobnostní modely, díky kterým je možno odhadovat pravděpodobnost $P(y|x)$, tedy pravděpodobnost přiřazení klasifikační skupiny (y) k vektoru příznaků (x) [20].

Pravděpodobnost P pro i -tý příznak lze odhadnout pomocí četnosti výskytu rysu v trénovacích datech ze vztahu 2.19

$$P(y|x_i) = \frac{P(x_i|y) \cdot P(y)}{P(x_i)}, \quad (2.19)$$

kde $P(y|x_i)$ lze intuitivně vysvětlit jako pravděpodobnost zařazení ke skupině y při výskytu určitého příznaku x_i s pravděpodobností $P(x)$ [20, 21].

Učení pravděpodobnostního modelu spočívá ve spočítání četností, které jsou potřebné pro výpočet pravděpodobností [20, 21].

Výhodou této metody je nízká časová náročnost, nicméně pro optimální fungování je nutné určit vhodné rozložení pravděpodobností. Zároveň je nutné extrahovat příznaky a v základních modelech volit kategoriální proměnné [20, 21].

2.3.2 Klasifikační stromy

Klasifikační stromy využívají učení s učitelem. Jedná se o struktury bez umělých neuronů, které bývají intuitivní a jednoduché na implementaci [18, 22].

Ve své podstatě se jedná o posloupnost rozhodnutí, jejímž výsledkem je zařazení prvku do jedné z klasifikačních skupin na základě vlastností prvku - konkrétních příznaků. Kořen stromu obsahuje celý datový soubor. Dále se skládá z větví, které se setkávají v uzlech, které jsou určeny proměnnými, podle kterých dělíme datový soubor a hranicemi, které určují, kde se dělení má provést. Každý list pak představuje jednotlivé klasifikační skupiny [18, 22].

Každý strom by měl být dostatečně přesný a zároveň i obecný, aby ho bylo možno aplikovat na různá data. Vhodné je použití křížové validace, při které jsou původní data rozdělena do k skupin, jedna skupina je použita jako testovací a zbytek pak pro tvorbu stromu. Validaci opakujeme tolikrát, dokud nedosáhneme optimálních výsledků [18, 22].

Data nemusí mít normální rozložení, zároveň mohou být vysoce korelovaná, nejsou na ně kladeny žádné požadavky [18, 22].

2.3.3 Metoda nejbližšího souseda

Metoda nejbližšího souseda (anglicky Nearest Neighbour, dále jen NN) slouží k zařazování objektů na základě třídy nejbližšího prvku se známou klasifikací - souseda. Tato metoda je náchylná k šumu (chyba v trénovací množině), proto je často modifikována [18, 23].

Vylepšená metoda vyhledává k nejbližších prvků z trénovací množiny se známou klasifikací a na základě těchto sousedů klasifikuje daný objekt. Tato modifikovaná metoda se nazývá k -NN [18, 23].

Pomocí vybrané metriky vzdálenosti se určí vzdálenost nezařazeného objektu od všech prvků v trénovací množině. Tento objekt je následně klasifikován do třídy, která se nejčastěji vyskytuje mezi k nejbližšími sousedy. Počet sousedů je libovolný, pomocí optimalizačních metod lze určit vhodný počet s nejlepšími výsledky a výpočetní náročností. Počet sousedů by mělo být každopádně liché číslo, pokud by byl počet sousedů sudý a došlo k nerozhodnosti při klasifikaci, výsledné přiřazení by bylo náhodné [18, 23].

U souborů dat s nevyváženými klasifikačními skupinami není použití této metody vhodné, jelikož je pravděpodobné, že bude daný objekt z méně zastoupené skupiny v datasetu obklopen prvky z více zastoupené skupiny, což by mohlo zkreslit výsledné zařazení objektu i při vyšším počtu sousedů. Výsledky k -NN klasifikace jsou závislé na zvolené metrice vzdáleností, hodnoty příznaků je třeba předpřipravit - normalizovat [18, 23].

2.3.4 Metoda podpůrných vektorů

Metoda podpůrných vektorů (anglicky Support Vector Machine, dále jen SVM) využívá strojového učení s učitelem. Tento klasifikační model má za úkol najít nejlepší hraniční nadrovinu, která rozdělí prostor příznaků tak, že trénovací data z různých klasifikačních skupin leží v opačných poloprostorech. V 2D prostoru tvoří hraniční nadrovinu křivka, v 3D prostoru rovina [26, 25].

Hraniční pásmo je oblast paralelní s hraniční nadrovinou bez bodů dat. Podpůrné vektory jsou body, které leží na okrajích hraničního pásma a slouží k popisu nadroviny. SVM klasifikuje data podle toho, ve kterém poloprostoru se nacházejí a to, jak daleko se nacházejí od hraniční nadroviny popisuje SVM skóre [26, 25].

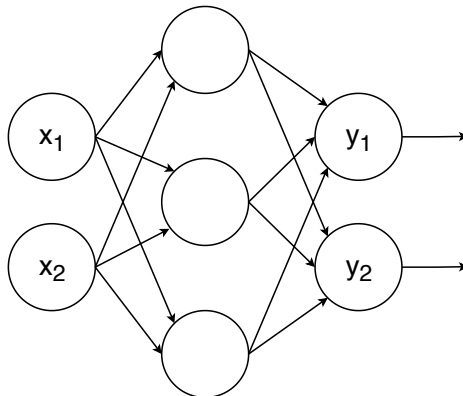
Modely SVM se dělí na lineární a nelineární. Při použití lineární metody nedochází k žádné jádrové transformaci a není změněna dimenzionalita původního příznakového prostoru. U nelineární metody SVM je potřeba zvolit jádrovou funkci, tzv. kernel function. Tato funkce umožní transformovat data do nového prostoru o vyšší dimenzi než původní prostor. V tomto novém prostoru se dále data opět rozdělují pomocí lineárních hranic a následně klasifikují [26, 25].

Pro využití SVM je třeba vhodně extrahovat a redukovat příznaky. K tomu se často využívá stacionární vlnková transformace, jejíž výkonové spektrum v každém stupni rozkladu signálu slouží k analýze síňové aktivity. Slouží také k získání parametrů peak-to-average ratio a log-energy entropie, na základě kterých lze klasifikovat jednotlivé úseky signálu [24].

Změna lineárního modelu na nelineární model patří mezi největší výhodu SVM, právě díky změně jádrové funkce (tzv. kernelizaci) je možné řešit nelineární problémy. Nevýhodou této metody je výpočetní náročnost při řešení úloh s rozsáhlými daty. [24, 26, 25].

2.3.5 Neuronové sítě

Nejpřesnější metody klasifikace představují neuronové sítě, kterých existuje mnoho typů. Neuronové sítě se skládají z umělých neuronů, tzv. perceptronů, které napodobují biologické funkce nervové buňky. Jsou to základní výpočetní jednotky, jsou mezi sebou propojeny spoji s váhami. Zpravidla výstup jednoho neuronu navazuje na další neuron. Mezi vstupními a výstupními neurony se může vyskytovat různý počet skrytých vrstev. Nejjednodušší neuronovou síť znázorňuje obrázek 2.5. x_1 a x_2 zde značí vrstvu se dvěma vstupními neurony, y_1 a y_2 vrstvu se dvěma výstupními neurony, mezi nimi se nachází jedna skrytá vrstva [29].



Obr. 2.5: Struktura neuronové sítě s jednou skrytou vrstvou.

Před samotným použitím neuronové sítě je zapotřebí připravit data pro učení. Důležitou částí je vybrání vhodné aktivační funkce, která je nezbytná pro zajištění konvergence výpočtu naučení neuronové sítě. Síť je třeba naučit, jak data klasifikovat, to lze provést s učitelem či bez učitele. Po kontrole chyb naučené sítě a optimalizaci lze přejít k výsledné klasifikaci a porovnání např. s expertními anotacemi [29].

3 Úvod do praktické části

Všechny algoritmy byly naprogramovány a testovány v prostředí Matlab R2020b.

3.1 Použitá data

Pro praktickou část této bakalářské práce byla využita data dostupná z databáze UBMI. Jedná se celkem o 300 jednosvodových EKG záznamů s vzorkovací frekvencí 300 Hz se známými anotacemi. Průměrná délka jednotlivých záznamů je 35 s. Signály byly pořízeny zařízením AliveCor a zastoupení jednotlivých srdečních rytmů v databázi je uvedeno v tabulce 3.1.

Tab. 3.1: Rozložení použitých dat

Název rytmu	Anotace	Počet záznamů v databázi
Fibrilace síní	A	100
Sinusový rytmus	N	100
Ostatní arytmie	O	100

Celý dataset byl dodán ve dvou samostatných podsetech, každý z nich obsahoval 150 signálů, které zahrnovaly 50 signálů s FIS, 50 fyziologických a 50 signálů s jinou arytmií. Toto původní rozdělení bylo zachováno, přičemž první část dat sloužila jako data trénovací, druhá část sloužila jako data testovací, na nichž byly otestovány vybrané metody. Toto rozdělení umožnilo otestovat algoritmy naprosto nezávisle na sadě dat, výsledky tak jsou naprosto nezkreslené trénovacími daty a lze je pokládat za relevantní.

K testování detektoru QRS komplexů byl využit tester poskytnutý Ústavem biomedicínského inženýrství Vysokého učení technického v Brně (UBMI) [30]. Celá databáze obsahovala 125 jednosvodových signálů, do testovací databáze bylo zahrnuto 123 signálů, dva signály nebyly zahrnuty kvůli kardiostimulátoru.

3.2 Parametry pro hodnocení úspěšnosti

Pro zhodnocení výsledků vybraných metod slouží několik parametrů, které lze vypočítat z hodnot v tzv. matici záměn (anglicky confusion matrix). Tato matice obsahuje 4 hodnoty, které srovnávají výsledky klasifikace do dvou kategorií. Obecnou matici záměn znázorňuje tabulka 3.2. Hodnota TP (anglicky true positive) znamená počet skutečně pozitivních výsledků, tedy kolik signálů s FIS bylo označeno jako patologický, TN (anglicky true negative) počet skutečně negativních výsledků, tedy počet

signálů bez FIS označených jako fyziologické. Hodnota FP (anglicky false positive) znamená, kolik výsledků bylo falešně pozitivních, tedy kolik signálů bez FIS bylo označeno jako patologické, FN (anglicky false negative) znamená počet falešně negativních výsledků, tedy kolik signálů s FIS bylo označeno jako fyziologické [27, 28].

Tab. 3.2: Obecná matice záměn.

Výsledná klasifikace	Referenční třída	
	Negativní	Pozitivní
Negativní	TN	FN
Pozitivní	FP	TP

Z hodnot TP, TN, FP a FN dále lze vypočítat parametry pro celkové zhodnocení funkčnosti implementovaného algoritmu. Mezi tyto parametry patří senzitivita, specificita, přesnost, pozitivní prediktivní hodnota, negativní prediktivní hodnota a f-skóre [27, 28].

Senzitivita (anglicky Sensitivity, dále jen SE) nabývá hodnot od nuly do jedné, případně od nuly do sta v procentech. Je to schopnost rozpoznat skutečně pozitivní (TP) objekty. Pro SE platí vztah 3.1 [27]

$$SE = \frac{TP}{TP + FN}. \quad (3.1)$$

Specificita (anglicky specificity, dále jen SP) nabývá stejně jako SE hodnot od nuly do jedné, případně od nuly do sta v procentech. Jedná se o schopnost rozeznat skutečně negativní (TN) objekty. Pro SP platí vzorec 3.2 [27]

$$SP = \frac{TN}{TN + FP}. \quad (3.2)$$

Přesnost (anglicky accuracy, dále jen ACC) vyjadřuje podíl správně klasifikovaných objektů ze všech testovaných objektů. Pro ACC platí vztah 3.3 [28]

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3.3)$$

Pozitivní prediktivní hodnota (anglicky positive predictive value, dále jen PPV) vyjadřuje pravděpodobnost, že pozitivně klasifikovaný objekt je skutečně pozitivní. Pro PPV platí vztah 3.4 [27]

$$PPV = \frac{TP}{TP + FP}. \quad (3.4)$$

Stejně jako PPV existuje i negativní prediktivní hodnota (anglicky negative predictive value, dále jen NPV), která pravděpodobnost, že negativně klasifikovaný objekt je skutečně negativní. Pro NPV platí vztah 3.5 [27]

$$NPV = \frac{TN}{TN + FN}. \quad (3.5)$$

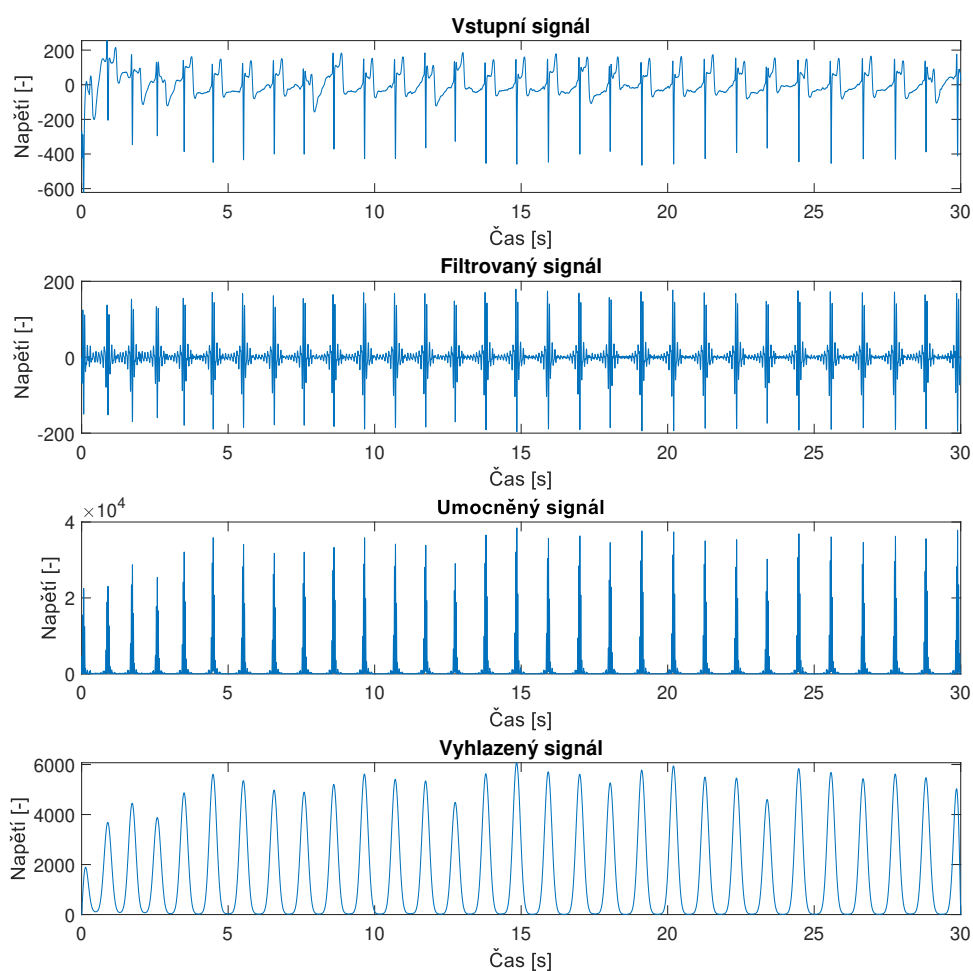
Parametr, který se používá v datasetech s nerovnoměrným rozdělením, se nazývá F-skóre. Pro F-skóre platí vztah 3.6 [27]

$$F - skóre = \frac{2 \cdot TP}{2 \cdot TP + FN + FP}. \quad (3.6)$$

4 Realizace detektoru QRS

4.1 Předzpracování dat

Každý EKG signál byl nejprve filtrován pásmovou propustí 10 Hz až 21 Hz pro zvýraznění QRS komplexů a potlačení ostatních složek signálu a rušení, např. síťového brumu nebo driftu. Následně byl signál pro zvýraznění oblastí komplexů QRS umocněn a vyhlazen dolní propustí s mezní frekvencí 2 Hz. Tyto kroky jsou zobrazeny na obrázku 4.1 a vedou k získání tzv. hrubého odhadu kvadrátu obálky filtrovaného signálu. Vstupním signálem je signál A00006 z datasetu trénovacích dat.



Obr. 4.1: Hrubý odhad kvadrátu obálky filtrovaného signálu.

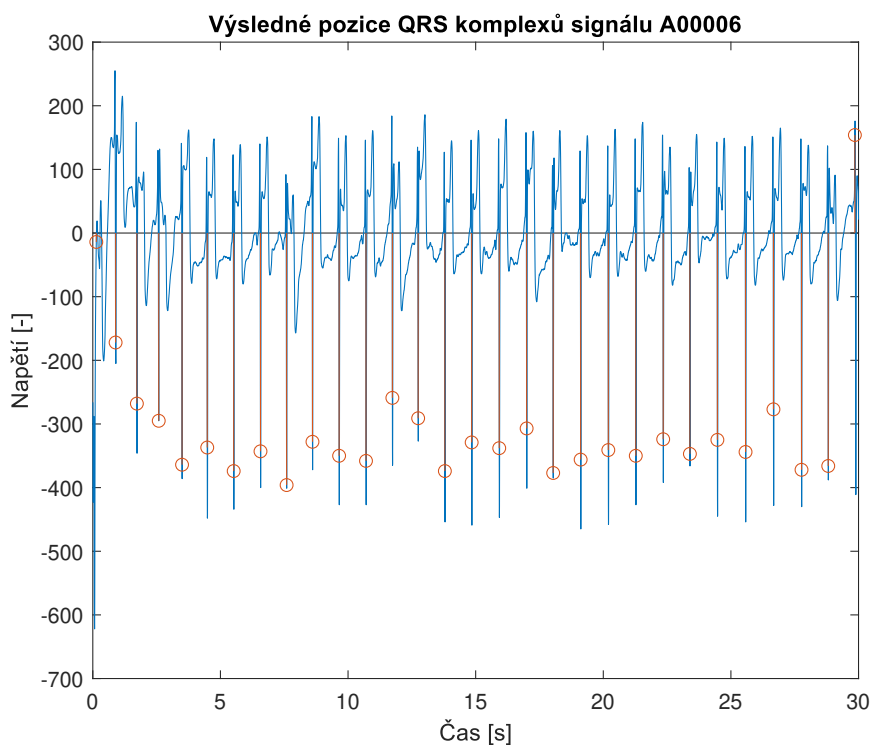
4.2 Rozhodovací pravidlo

Pro určení pozic QRS komplexů bylo následně použito rozhodování na základě překročení adaptivního prahu. Na začátku je práh určen z počátečních dvou vteřin filtrovaného signálu jako 40 % maximální hodnoty. Každé dvě další vteřiny je práh obdobně adaptován. Signál je procházen po prvcích a při překročení prahu je vyhledáváno maximum. Pokud dojde k detekci, je uložena jeho pozice a dojde k posunu o refrakterní fázi 0,15 vteřin. Během této doby není očekáván žádný další QRS komplex. Pokud není po 1,66násobku délky předchozího RR-intervalu nalezen žádný QRS komplex, práh se sníží na 40 % původního prahu a hledání probíhá zpětně.

4.3 Výsledky detekce QRS komplexů

Tento algoritmus byl otestován CSE QRS Testerem dostupného z ÚBMI [30]. Tento software slouží k výpočtu senzitivity a pozitivní prediktivní hodnoty algoritmu a k výpočtu časové náročnosti algoritmu, která spíše poukazuje na kvalitu použitého počítače, než samotného algoritmu.

Na obrázku 4.2 je znázorněn signál A00006 z trénovacího datasetu se zvýrazněnými pozicemi QRS komplexů tak, jak je detekoval použitý algoritmus.



Obr. 4.2: Signál A00006 s vyznačenými QRS komplexy.

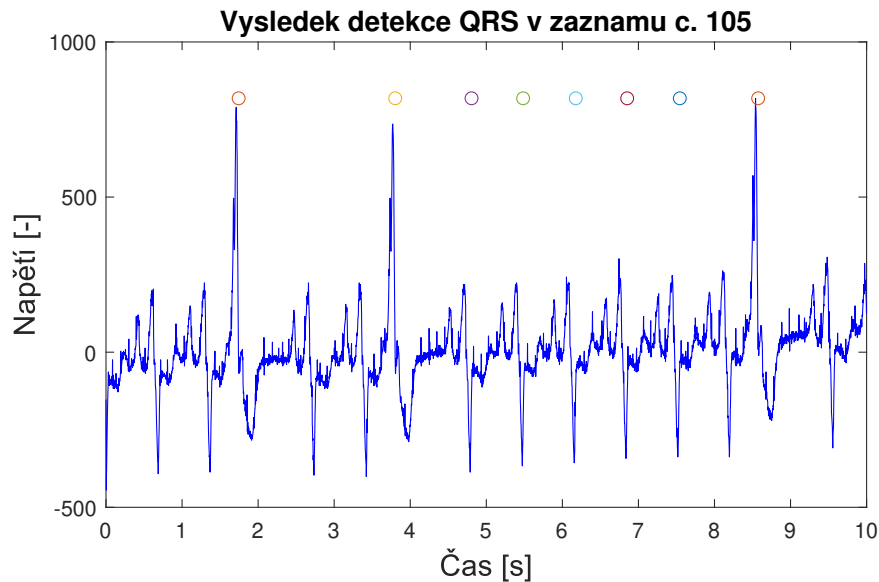
Výsledky testování algoritmu jsou uvedené v tabulce 4.1. Senzitivita algoritmu byla 97,56 %, pozitivní prediktivní hodnota 99,58 %. Průměrný čas výpočtu byl 1,1 vteřin.

Tab. 4.1: Výsledky testování QRS detektoru.

Parametr	Hodnota
SE	97,56 %
PPV	99,58 %

4.4 Diskuze výsledků detekce QRS komplexů

Zvoleným postupem se podařilo detekovat QRS komplexy se senzitivitou 97,56 % a s pozitivní prediktivní hodnotou 99,58 %. Algoritmus využíval rozhodovací pravidlo na základě překročení adaptivního prahu. Zvolený práh nemusel být vždy optimální. Na obrázku 4.3 je zobrazen signál č. 105 z CSE databáze pro testování QRS komplexů, u kterého zvolený algoritmus detekoval osm ze čtrnácti QRS komplexů. V prvních dvou vteřinách je práh nastaven na 40 % z maximální abnormální výchylky, proto nebyly první dva QRS komplexy detekovány.



Obr. 4.3: Signál 105 se špatně detekovanými QRS komplexy.

5 Výběr příznaků pro použití vybraných metod

Na začátku bylo vybráno deset příznaků vycházejících z nepravidelnosti RR intervalů. Konkrétně se jednalo o příznaky:

1. CV - Kovarianční koeficient, viz kapitola 2.2.1.
2. TPR - Počet zlomových bodů, viz kapitola 2.2.8.
3. pNN20 - Hodnota NN20 vyjadřuje počet párů sousedních RR intervalů, které se od sebe liší více než o 20 ms. Příznak pNN20 je poměr NN20 a celkového počtu RR intervalů.
4. pNN50 - Hodnota NN50 vyjadřuje počet párů sousedních RR intervalů, které se od sebe liší více než o 50 ms. Příznak pNN50 je poměr NN50 a celkového počtu RR intervalů.
5. MAD - Medián absolutní odchylky od střední hodnoty, viz kapitola 2.2.4.
6. SE - Vyjadřuje entropii dat, viz kapitola 2.2.7.
7. nRMSSD - Normalizovaná střední kvadratická odchylka, viz kapitola 2.2.2.
8. SD1/SD2 - Poměr rychlé variability ku dlouhodobé variabilitě srdečního rytmu, viz kapitola 2.2.10.
9. Koeficient šikmosti - Popisuje rozložení dat, viz kapitola 2.2.5.
10. Koeficient špičatosti - Popisuje rozložení dat kolem střední hodnoty, viz kapitola 2.2.6.

Byly vytvořeny jednotlivé příznaky pro signály s fibrilacemi, pro signály s ostatními arytmiemi i bez arytmií. Příznaky pro skupiny s fibrilacemi a pro signály s normálním sinusovým rytmem byly podrobeny statistické analýze. Ke statistickému zhodnocení příznaků byl využit software STATISTICA verze 13.5. Cílem bylo zjistit, které z 10 naprogramovaných příznaků vyjmenovaných výše vhodně rozlišují signály bez a s fibrilacemi.

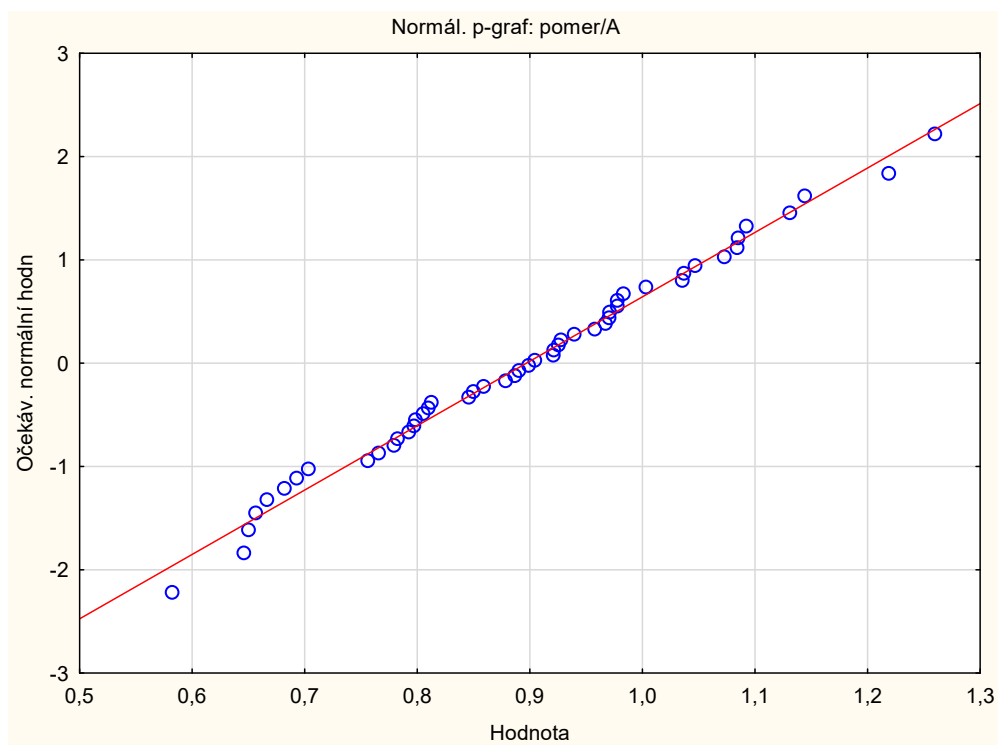
V jednotlivých testech v softwaru STATISTICA 13.5 je k hodnotě testové statistiky připojována vypočítaná tzv. p-hodnota, která představuje hladinu významnosti, na jejímž základě je možno přijmout či zamítnout nulovou hypotézu. Nulovou hypotézu zamítáme, pokud je p-hodnota menší než zvolená hodnota alfa, která je nejčastěji stanovena na hodnotu 0,05 (5 %). Tato hodnota značí pravděpodobnost chyby prvního druhu, která spočívá v chybném zamítnutí pravdivé nulové hypotézy. Pokud je p-hodnota větší, než hladina významnosti, nulovou hypotézu nezamítáme [31].

5.1 Testy normality

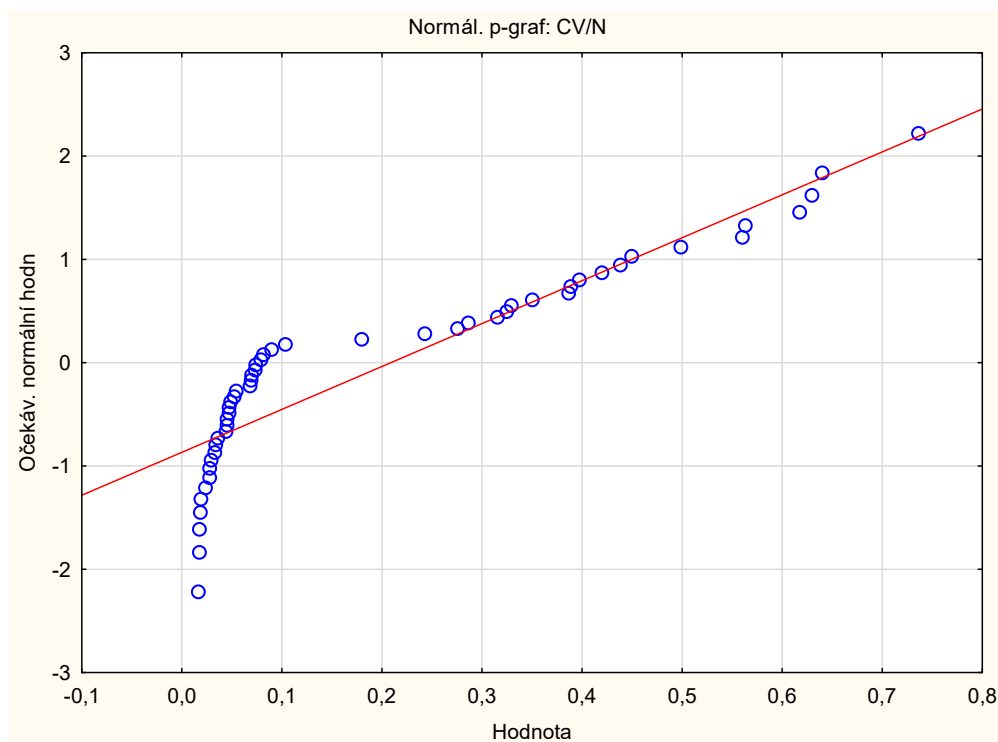
Příznaky pro signály s fibrilacemi a pro signály bez fibrilací byly podrobeny analýze normálního rozložení dat. Příznaky pro signály s ostatními arytmiemi nebyly analyzovány vzhledem k tomu, že cílem bylo zjistit, zda příznaky vhodně rozlišují signály s fibrilacemi a bez fibrilací.

Normalita dat byla testována pomocí Shapiro-Wilkova testu. Nulovou hypotézou bylo, že data vykazují normální rozložení dat. Hladina významnosti byla zvolena jako hodnota 0,05. Pokud by p-hodnota byla menší, než zvolená hladina významnosti, data nemají normální rozložení.

Pro ověření bylo využito grafických metod, konkrétně vykreslení pravděpodobnostních grafů, ukázkou grafu pro příznak s normálním rozložením příznaku poměru $SD1/SD2$ pro signály s fibrilacemi lze vidět na obrázku 5.1 a graf pro příznak CV signálů bez fibrilací s nenormálním rozložením lze vidět na obrázku 5.2.



Obr. 5.1: P-graf zobrazující normální rozložení.



Obr. 5.2: P-graf zobrazující nenormální rozložení.

5.2 Testy statistické významnosti

Z příznaků pro signály bez fibrilací normální rozložení dat vykazovaly pouze příznaky pNN20 a poměr SD1/SD2. Z příznaků pro signály s fibrilacemi vykazoval normální rozložení dat pouze příznak poměr SD1/SD2.

K testování statistické významnosti byly využity dva testy, konkrétně Mann-Whitneyův U test a nepárový t-test.

Mann-Whitneyův U test je neparametrický test, který se používá pro hodnocení nepárových pokusů, kdy porovnáváme 2 různé výběrové soubory (pokusný zásah A, B). Testujeme hypotézu, že veličina X odpovídající pokusnému zásahu „A“ a veličina Y odpovídající pokusnému zásahu „B“ mají totéž rozdělení pravděpodobností. Tento test je vhodný pro veličiny, které nemusí odpovídat Gaussovu normálnímu rozdělení [32].

Nepárový t-test slouží pro porovnání dat z dvou nezávislých výběrů, tzn. že pocházejí ze dvou různých skupin jedinců. Test slouží pro porovnání středních hodnot, přičemž nulová hypotéza tvrdí, že střední hodnoty obou výběrů jsou si rovny. Testuje data s Gaussovým normálním rozdělením [33].

Všechny příznaky byly testovány vybraným testem na hladině významnosti 0,05 a výsledky jsou uvedeny v tabulce 5.1.

Tab. 5.1: Výsledky statistických testů.

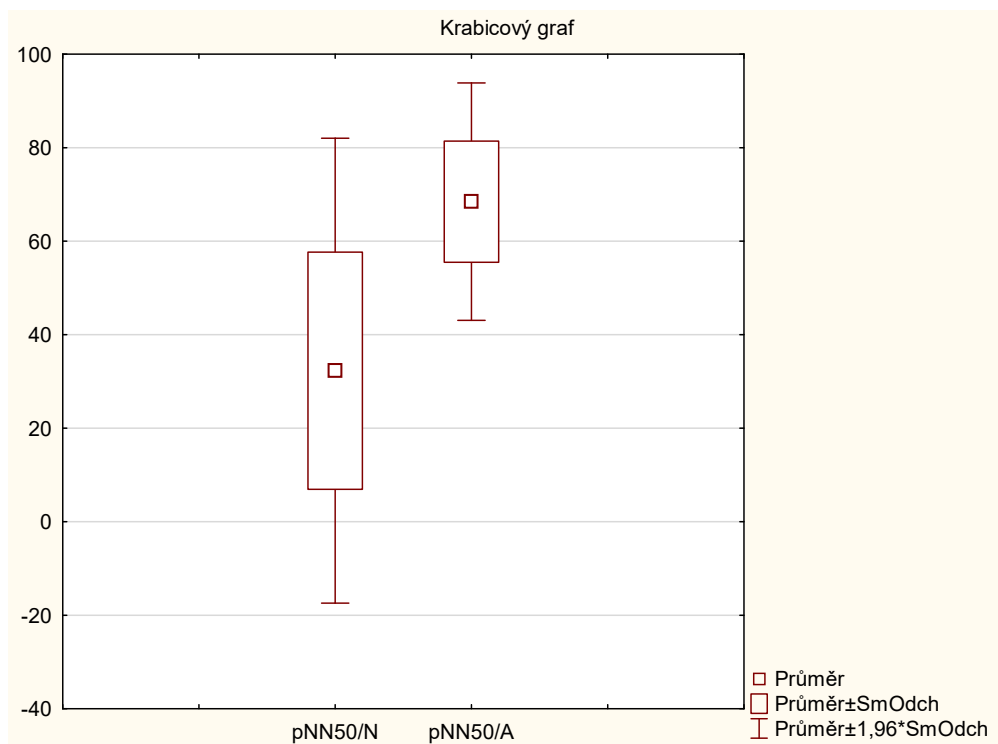
VÝSLEDKY STATISTICKÝCH TESTŮ		
Příznak	p-hodnota	Statisticky významné
CV	0,000301	ANO
TPR	0,000010	ANO
pNN20	0	ANO
pNN50	0	ANO
nRMSSD	0	ANO
MAD	0,000001	ANO
SE	0,037036	ANO
poměr SD1/SD2	0,002421	ANO
špičatost	0,942295	NE
šikmost	0,641693	NE

Z tabulky výsledků statistických testů 5.1 lze vyčíst, že prvních 8 příznaků mělo p-hodnotu menší, než byla zvolená hladina významnosti, což znamená, že byla zamítnuta nulová hypotéza o shodě středních hodnot, popřípadě o stejném pravděpodobnostním rozložení a tyto příznaky byly ponechány pro klasifikaci, dokázaly totiž rozlišit data bez a s fibrilacemi. Koeficienty špičatosti a šikmosti měly výslednou p-hodnotu větší, než kritickou hodnotu, proto nebyla zamítnuta nulová hypotéza a tyto příznaky byly vyřazeny jako statisticky nevýznamné.

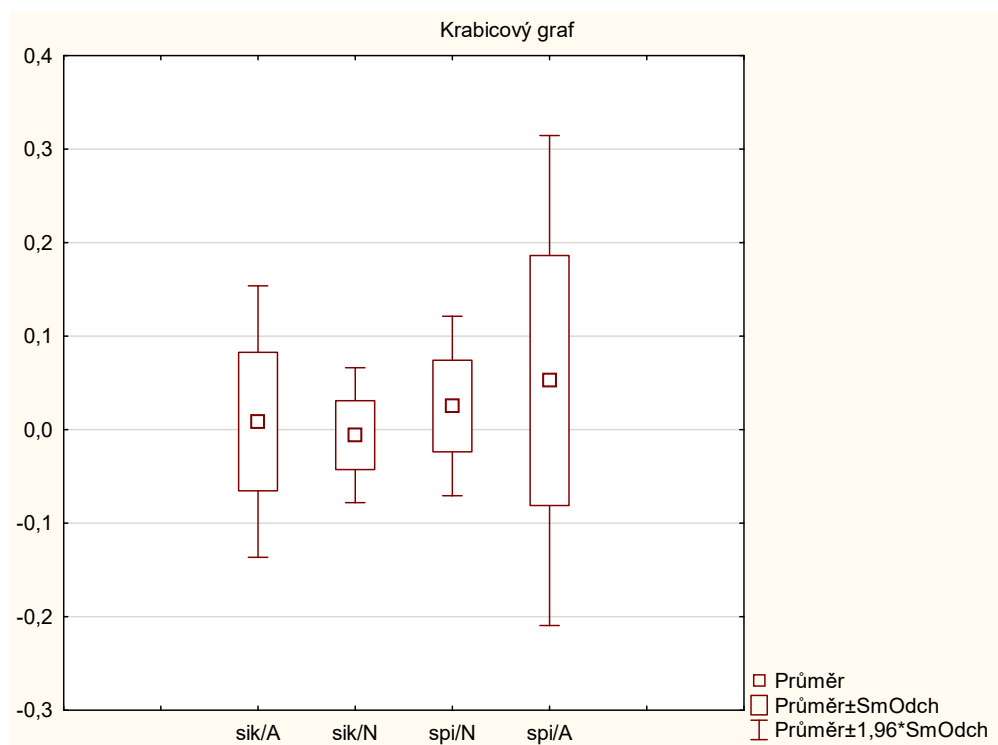
Na obrázku 5.3 je zobrazen krabicový graf statisticky významného příznaku pNN50 a na obrázku 5.4 jsou zobrazeny krabicové grafy pro statisticky nevýznamné koeficienty špičatosti a šikmosti. Označení N za lomítkem v grafu popisuje skupinu bez fibrilací, označení A za lomítkem popisuje skupinu s fibrilacemi.

U statisticky významného příznaku pNN50 je z grafu patrná vyšší průměrná hodnota. Na druhém grafu jsou průměrné hodnoty velice blízké a příznaky nerozlišují dobře skupiny bez a s fibrilací.

Z počátečních deseti příznaků tak byl počet příznaků zredukován na výsledných osm, které byly dále využity v jednotlivých vybraných metodách.



Obr. 5.3: Krabicový graf pro statisticky významný příznak pNN50.



Obr. 5.4: Krabicový graf pro statisticky nevýznamné příznaky.

6 Použití prahovacích metod

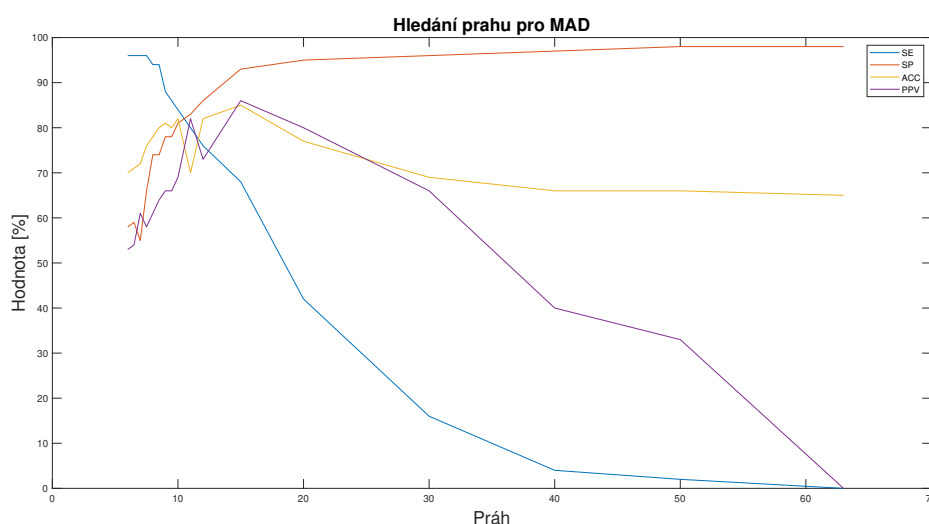
Po detekování pozic R vln a zjištění délek RR intervalů byly vyřazeny celkem čtyři ektopické cykly, dva nejdelší RR intervaly a dva nejkratší RR intervaly.

Vstupem do algoritmu vyhodnocujícího FIS byl jeden signál EKG. Ten byl vyhodnocován na základně vybraných příznaků, nakonec byl použit většinový hlasovací systém. Výstupem algoritmu je hodnota 0 nebo 1, přičemž 0 značí signál bez FIS (anotace N - normální sinusový rytmus či O - ostatní arytmie), zatímco hodnota 1 značí signál s FIS. Tato hodnota se následně porovnává s expertní anotací a připsuje se do matice záměn, viz tabulka 3.2. Ze všech hodnot jsou následně podle vztahů uvedených v kapitole 3.2 vypočítány parametry SE, SP, ACC a PPV.

6.1 Prahy pro jednotlivé příznaky

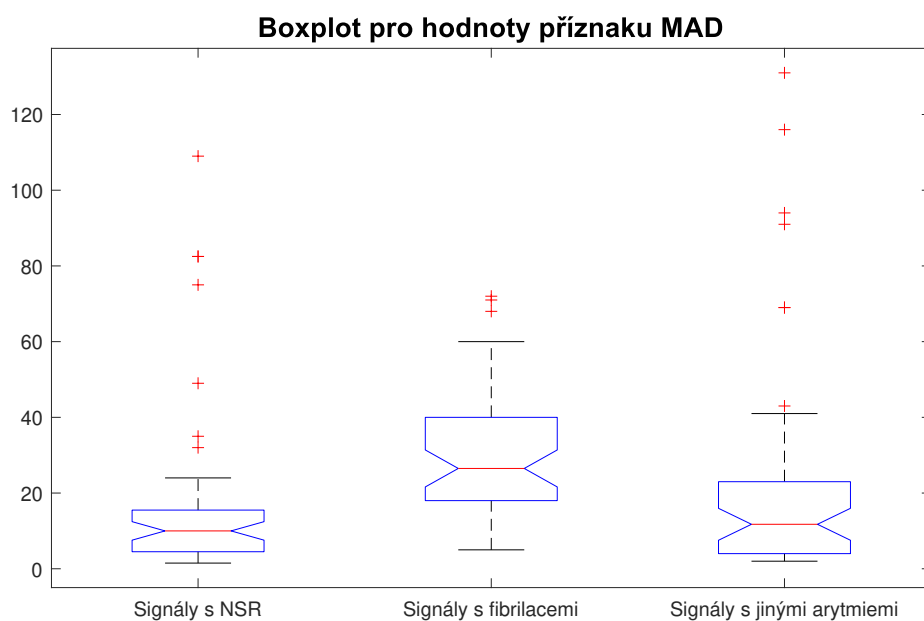
Pro každý příznak byl zvolen optimální práh pro zachování co nejlepších výsledků u trénovacích dat.

Na začátku byla pro každý příznak zjištěna minimální a maximální hodnota. V rozmezí těchto hodnot pak s dostatečně malým krokem probíhalo počítání hodnot SE, SP, ACC a PPV. Z těchto hodnot byl následně vytvořen graf, který je možno vidět i na obrázku 6.1. Z tohoto obrázku je patrné, že při nízkých hodnotách prahu se nedařilo odlišit signály s arytmiemi a bez, při příliš vysokém prahu hodnoty SE a PPV strmě klesaly. Obecně se zvyšováním prahu snižovala senzitivita a pozitivní prediktivní hodnota, specificita rostla, protože více signálů bylo označeno jako bez arytmie, díky tomu byly v podstatě všechny patologické signály nalezeny.



Obr. 6.1: Křivky parametrů při optimalizaci prahu.

Tímto způsobem byl vyhodnocen každý z příznaků a vybrán optimální práh. Tyto prahy pak již dále nebyly měněny, je tedy možné, že v kombinaci tyto prahy nebyly úplně optimální. Pro MAD byl nakonec vybrán práh s hodnotou 10. Pro ilustraci je na obrázku 6.2 vyobrazen krabicový graf pro hodnoty příznaků MAD pro signály s normálním sinusovým rytmem, signály s fibrilacemi a signály s jinými arytmiemi. Na tomto boxplotu lze vidět, že rozsah hodnot u signálů s arytmiemi je mnohem větší, MAD zde dosahuje vyšších hodnot, je také větší průměrná hodnota.

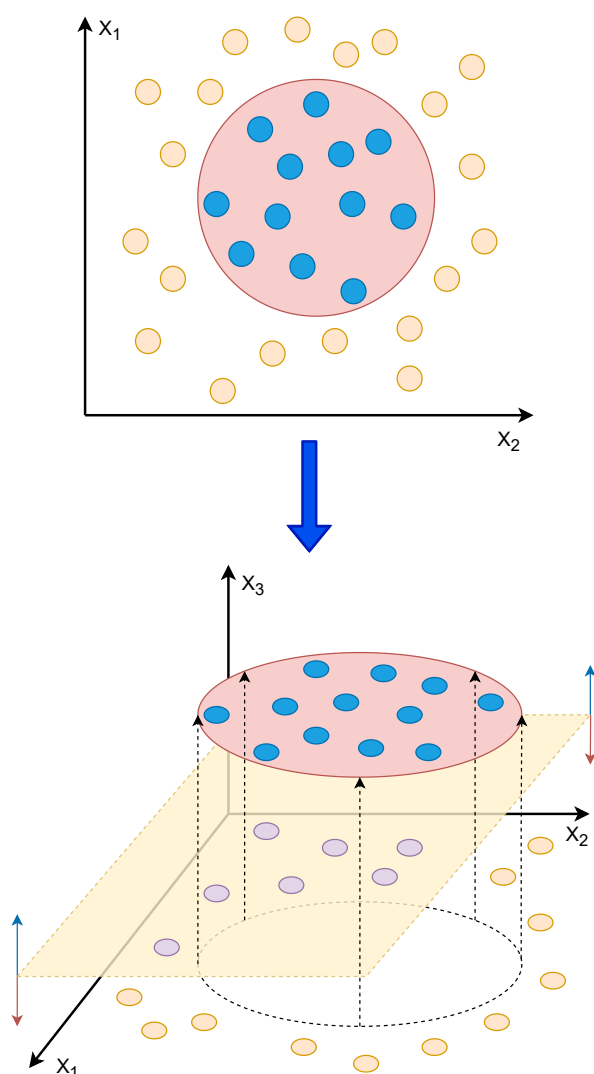


Obr. 6.2: Hodnoty MAD pro tři typy signálů.

7 Použití SVM klasifikátoru

Metoda podpůrných vektorů byla popsána v kapitole 2.3.4. K trénování i klasifikaci dat byly využity implementované funkce prostředí MATLAB R2020b, které umožňují vytvořit modely s různými typy jader, nastavení modelů probíhá přímo v rámci funkce.

Metoda SVM byla vybrána především pro možnost řešení nelineárních problémů. Na obrázku 7.1 převzatého z [35] je zobrazeno, jakým způsobem probíhá řešení nelineárních problémů. V horní části lze vidět dvě klasifikační skupiny znázorněné modrými a červenými kolečky, které nelze lineárně oddělit ve 2D prostoru. Pokud dojde k umělému zvýšení dimenzí a přesunu příznaků do 3D prostoru, skupiny lze vhodně klasifikovat.



Obr. 7.1: Princip vzniku lineárního řešení nelineárního problému [35].

7.1 Vytvoření klasifikátoru

K trénování i testování byly využity stejné sady dat, jako v případě prahovacích metod. Prvním krokem bylo vytvoření klasifikátoru pomocí funkce *fitcsvm*. Základními vstupy byly příznaky odvozené z trénovacích dat a klasifikační skupiny k příslušným signálům z trénovací množiny. Jednotlivé řádky v buňkovém poli značí jeden objekt z trénovací množiny. Anotace klasifikátoru rozlišují dvě skupiny signálů, první skupina obsahuje signály s fibrilacemi, ta nese označení 1, skupina nesoucí označení 0 obsahuje signály bez arytmií nebo s jinými arytmiemi, než jsou fibrilace.

V rámci funkce je možné nastavit parametr *KernelFunction*. Tento parametr, jak již z názvu napovídá, určuje typ jádra funkce. Bylo vyzkoušeno několik typů jader, od lineárního po polynomiální a byly zaznamenány výsledky klasifikace testovacích dat.

7.2 Klasifikace dat

Samotná klasifikace dat probíhá pomocí příkazu *predict*. Do této funkce vstupuje náš SVM klasifikátor a testovací data. Výsledkem jsou jednotlivé klasifikační skupiny pro každý signál z testovacího datasetu a skóre, s jakým je daný EKG záznam zařazen do třídy 0 nebo 1.

Výsledné klasifikační skupiny testovacích dat byly porovnány s expertními anotacemi, které byly k záznamům dodány a následně byly vypočítány hodnoty SE, SP, ACC a PPV.

8 Použití metody kNN

Metoda k-NN byla popsána v kapitole 2.3.3. Pro klasifikaci dat byla tato metoda zvolena z důvodu poměrně snadné implementace v prostředí MATLAB R2020b. Tato metoda je vhodná pro použití především kvůli rovnoměrnému rozložení dat v klasifikačních skupinách, proto by nemělo dojít k tomu, že prvek z množiny jedné skupiny bude obklopen převážně prvky jiné z klasifikační skupiny, která by tak měla početní převahu a došlo by ke špatnému zařazení prvku. Jako trénovací i testovací data byly použity stejné sady dat jako v případě použití prahovacích metod.

8.1 Vytvoření klasifikátoru

V prostředí MATLAB R2020b je možné k vytvoření a následnému použití klasifikátoru kNN využít již implementované funkce.

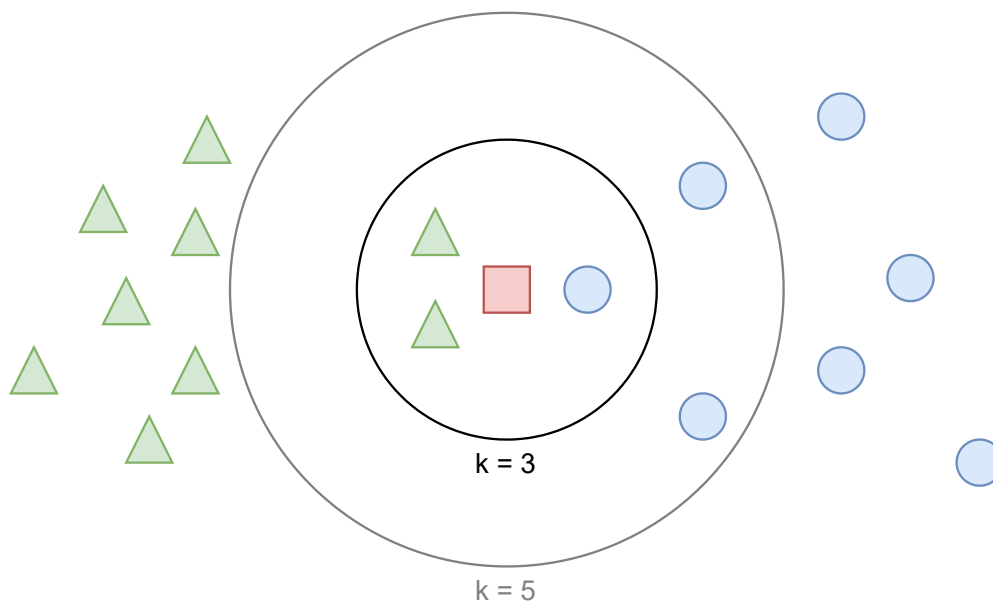
Pro vytvoření klasifikátoru slouží funkce *fitcknn*. Základními vstupy jsou trénovací data, respektive příznaky z nich odvozené v matici, přičemž objekty jsou zapsány v řádcích a příznaky ve sloupcích, dále jednotlivé klasifikační skupiny a počet sousedů. V rámci funkce se dále dají nastavit i další parametry, jako například metrika vzdálenosti, jako výchozí je ve funkci nastavená euklidovská vzdálenost. Přímo ve funkci lze data standardizovat, není třeba data upravovat před vytvořením klasifikátoru.

8.2 Výběr počtu sousedů

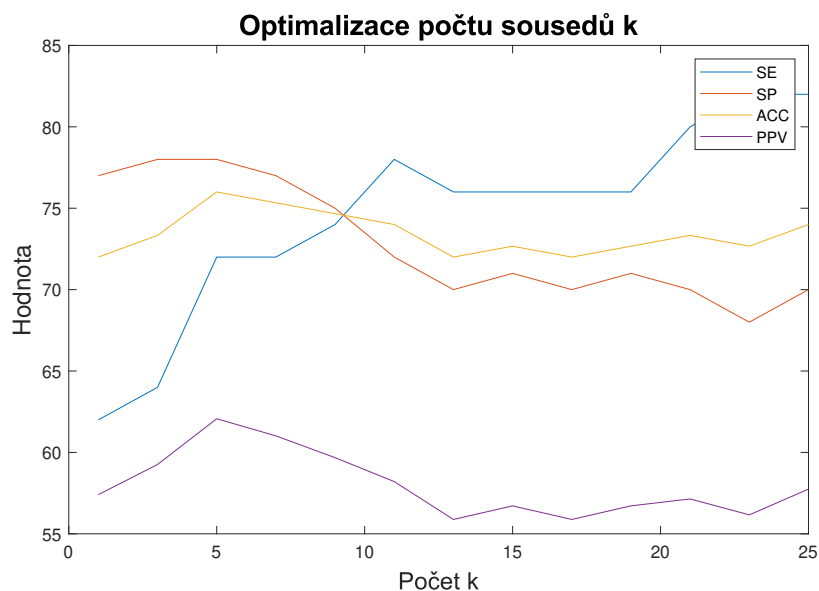
Počet sousedů k je rozhodující pro zařazení prvku. To, jak může počet sousedů ovlivnit zařazení prvku je vidět na obrázku 8.1. Na obrázku lze vidět dvě klasifikační skupiny - zelené trojúhelníky a modrá kolečka - a prvek, který je určen ke klasifikaci - červený čtverec. Pokud bude zvoleno $k = 3$, neznámý prvek bude přiřazen k zeleným trojúhelníkům, protože mají početní převahu ve vzdálenostech tří prvků od neznámého prvku. Pokud bude $k = 5$, prvek bude přiřazen k modrým kolečkům. Je patrné, že pokud by byl zvolen sudý počet sousedů, například dva, mohlo by dojít k nerozhodnosti v klasifikaci mezi modrým kolečkem a zeleným trojúhelníkem, bylo by nutné použít další rozhodovací pravidlo. Jak již bylo zmíněno v kapitole 2.3.3, počet sousedů k je lepší volit jako liché číslo než sudé, protože nemůže dojít k nerozhodnosti při klasifikaci.

Na začátku byl počet sousedů volen od 1 s krokem 2 až po hodnotu 50, přičemž v každém cyklu byly vypočítány hodnoty SE, SP, PPV a ACC. Z grafu pro tyto hodnoty bylo zjištěno, zhruba od hodnoty 25 se ukazatele úspěšnosti příliš neměnily. Byl proto proveden druhý cyklus, ve kterém se počet sousedů opět zvyšoval od

1 s krokem 2 po hodnotu 25 a z hodnot ukazatelů úspěšnosti byl vykreslen graf, který je vidět na obrázku 8.2. Z tohoto grafu a z matice hodnot pro jednotlivé počty sousedů bylo zjištěno, že nejlepší výsledky měl klasifikátor s 9 a s 11 sousedy. Při počtu sousedů přes 21 sice byla senzitivita nejnižší, jiné hodnoty však poklesly pod hodnoty při 9 sousedech.



Obr. 8.1: Závislost výsledné klasifikace na počtu sousedů.



Obr. 8.2: Optimalizace počtu sousedů.

8.2.1 Křížová validace

Pro ověření, že tyto počty sousedů nejsou příliš uzpůsobeny testovací sadě dat, byla provedena trojnásobná křížová validace. V rámci funkce *fitcknn* je možné povolit křížovou validaci, funkce zahrnuje desetinásobnou křížovou validaci, ta je vhodná pro rozsáhlejší sady dat, v tomto případě je vhodnější trojnásobná nebo pětinasobná křížová validace.

Všechna data (trénovací i testovací) byla sloučena dohromady, dohromady tak bylo k dispozici 300 signálů. Tato data byla rozdělena na třetiny s co nejrovnoměrnějším zastoupením klasifikačních skupin. Pro trénování byly použity vždy dvě třetiny dat, pro testování pak zbylá jedna třetina dat. Tímto způsobem byla provedena trojnásobná křížová validace.

Na trénovacích datech byl vždy vytvořen klasifikátor s počtem sousedů zvyšujícím se od 1 s krokem 2 po hodnotu 25 a tento klasifikátor byl použit pro validaci testovacích dat.

Při křížové validaci bylo zjištěno, že počet sousedů 9 a 11 jsou opravdu nejvhodnějšími počty sousedů pro klasifikaci této sady dat. Výsledky těchto dvou klasifikátorů se příliš nelišily, byl proto empiricky vybrán klasifikátor s jedenácti sousedy.

8.3 Klasifikace dat

Klasifikace probíhá stejně jako u metody SVM pomocí příkazu *predict*. Do této funkce vstupuje navržený kNN klasifikátor a testovací data. Výsledkem jsou klasifikační skupiny pro jednotlivé záznamy z testovací databáze, které jsou následně porovnávány s expertními anotacemi a jsou vypočítané hodnoty parametrů SE, SP, ACC a PPV.

9 Dosažené výsledky

9.1 Výsledky prahovacích metod

9.1.1 Detekce trénovacích dat

Ze všeho nejdříve byla otestována trénovací sada dat, na začátku proběhlo testování se všemi příznaky, poté byly vyzkoušeny různé kombinace příznaků pro maximalizování hodnot SE, SP, PPV a ACC. Ze všech pokusů nakonec nejlépe dopadla kombinace všech příznaků, tento výsledek byl zapsán pro přehlednost do tabulky 9.1. Při ubírání příznaků ve většině případů rostla senzitivita a klesala specificita a vznikalo mnoho falešně negativních případů, z tohoto důvodu nakonec byla vybrána kombinace všech příznaků, při které nebyla upřednostněna hodnota žádného parametru.

Tab. 9.1: Výsledky prahovacích metod pro trénovací data.

VÝSLEDKY PRO TRÉNOVACÍ DATA				
Zvolené příznaky	SE [%]	SP [%]	ACC [%]	PPV [%]
Všechny	84	66	73	57
pNN20, pNN50, MAD, SE, nRMSSD	92	89	90	81

Po porovnání anotací pro jednotlivé signály s referencemi dodaným k signálům byla vytvořena matice záměn pro trénovací data, která je v tabulce 9.2. Tato tabulka byla vytvořena po vyhodnocení trénovacích dat prahováním pomocí pěti vybraných příznaků s prahy optimalizovanými pro tuto kombinaci. Z této tabulky lze vyčíst, že na u trénovacích dat došlo ke špatné klasifikaci 3 signálů, které byly nesprávně vyhodnoceny jako falešně negativní, 15 signálů bylo vyhodnoceno jako falešně pozitivních.

Tab. 9.2: Matice záměn pro trénovací data pro prahovací metody.

Výsledná klasifikace	Referenční třída	
	Negativní	Pozitivní
Negativní	85	3
Pozitivní	15	47

9.1.2 Detekce testovacích dat

Stejný postup byl aplikován i na testovacím datasetu. Výsledky na testovacích datech jsou naprosto nezávislé, jelikož prahy byly optimalizovány pro trénovací data. To je

i důvodem, proč byl už dopředu očekáván pokles hodnot SE, SP, ACC a PPV. Prahy jsou příliš uzpůsobeny pro trénovací data. Výsledky testování algoritmu na testovacích datech jsou uvedeny v tabulce 9.3.

Tab. 9.3: Výsledky prahovacích metod pro testovací data.

VÝSLEDKY PRO TESTOVACÍ DATA				
Zvolené příznaky	SE [%]	SP [%]	ACC [%]	PPV [%]
Všechny	80	62	68	51
pNN20, pNN50, MAD, SE, nRMSSD	86	75	79	63

Tabulka 9.2 je maticí záměn pro testovací data. Z tabulky je patrné, že na testovacích datech bylo špatně vyhodnoceno 18 záznamů, které byly označeny za falešně pozitivní a 27 signálů bylo vyhodnoceno jako falešně negativní.

Tab. 9.4: Matice záměn pro testovací data pro prahovací metody.

	Referenční třída	
Výsledná klasifikace	Negativní	Pozitivní
Negativní	77	7
Pozitivní	23	43

9.2 Výsledky SVM klasifikace

Testovací data byla klasifikována SVM klasifikátorem natrénovaným na trénovacích datech. Výsledné klasifikační skupiny byly porovnány s dodanými referencemi a byly vypočítány hodnoty SE, SP, ACC a PPV pro klasifikátory s různými typy jader. Výsledky byly zaznamenány do tabulky 9.5.

Tab. 9.5: Výsledky metody SVM na testovacích datech.

SVM KLASIFIKACE TESTOVACÍCH DAT				
Typ jádra	SE [%]	SP [%]	ACC [%]	PPV [%]
Lineární	68	73	71	56
Gaussovské	56	88	77	70
Polynomiální st. 2	78	75	76	61
Polynomiální st. 3	74	75	74	60
Polynomiální st. 4	62	82	75	63

Z tabulky je patrné, že typ jádra výsledky příliš neměnil. Výsledky byly vesměs podobné, jako nejlepší je možno empiricky označit rozdělení klasifikátorem s polynomiálním jádrem druhého stupně. Při zvyšování polynomiálního stupně klesala senzitivita, zatímco specifita rostla až k hodnotě přes 90 %, pravděpodobně kvůli tomu, že narůstal počet signálů klasifikovaných do skupiny 0, tedy bez jakékoli arytmie.

Pro klasifikátor s polynomiálním jádrem druhého stupně byla vytvořena matice záměn, jejíž hodnoty jsou uvedeny v tabulce 9.6.

Tab. 9.6: Matice záměn pro testovací data pro metodu SVM.

Výsledná klasifikace	Referenční třída	
	Negativní	Pozitivní
Negativní	76	11
Pozitivní	24	39

9.3 Výsledky KNN klasifikace

Klasifikátor byl optimalizován pomocí trojnásobné křížové validace. Výsledky byly graficky zaznamenány pro lepší přehlednost, grafy závislostí hodnot statistických parametrů pro jednotlivé křížové validace jsou na obrázku 9.1.

V kapitole 8 již bylo zmíněno, že byl empiricky vybrán podle nejlepších výsledků klasifikátor s jedenácti sousedy. Na základě křížových validací bylo zjištěno, že počet sousedů není přizpůsoben pouze testovací sadě, byla tedy ověřena generalizační schopnost natrénovaného KNN modelu. Pro nejlepší klasifikátor s $k = 11$ byla vytvořena tabulka 9.7 s výsledky bez křížové validace i výsledky jednotlivých křížových validací.

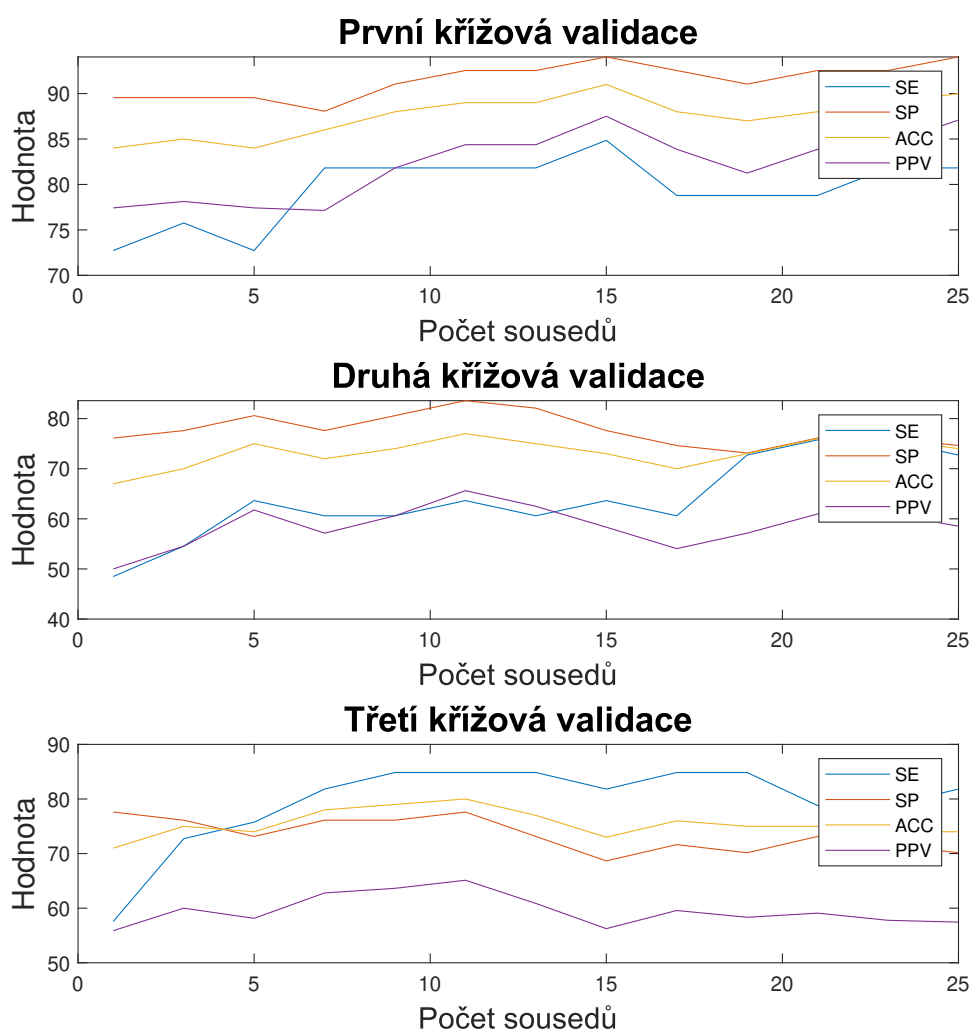
Tab. 9.7: Výsledky na testovacích datech pro klasifikátor s $k = 11$.

KNN KLASIFIKACE TESTOVACÍCH DAT				
Typ jádra	SE [%]	SP [%]	ACC [%]	PPV [%]
Bez validace	78	72	74	58
VÝSLEDKY KŘÍŽOVÝCH VALIDACÍ				
1. Křížová validace	79	92	88	84
2. Křížová validace	64	84	77	66
3. Křížová validace	85	76	80	65

Tabulka 9.8 maticí záměn pro klasifikaci testovacích dat. Z matice je patrné, že v porovnání s prahovacími metodami vzniklo méně falešně pozitivních, na druhou stranu došlo k nárůstu falešně negativních případů.

Tab. 9.8: Matice záměn pro testovací data pro metodu KNN.

Výsledná klasifikace	Referenční třída	
	Negativní	Pozitivní
Negativní	72	11
Pozitivní	28	39



Obr. 9.1: Grafy závislosti hodnot statistických parametrů pro jednotlivé křížové validace.

9.4 Vyřazení ostatních arytmií ze souboru dat

V mnoha studiích, ze kterých byla provedena literární rešerše, autoři používali jako trénovací a testovací data signály pouze s fibrilacemi nebo bez, do klasifikace nebyly zahrnuty signály s ostatními arytmiemi. Pro zjištění, jak moc jsou výsledky ovlivněny signály s ostatními arytmiemi byly tyto záznamy z testovací i trénovací množiny vyřazeny a všechny metody byly vyzkoušeny znovu se stejnými parametry na upravených datech. Byly využity stejné trénovací i testovací množiny, pouze bez signálů skupiny O, tedy s jinými poruchami srdečního rytmu.

Bylo očekáváno, že dojde ke zlepšení výsledků, protože ne vždy se dařilo dobře odlišit ostatní arytmie od fibrilací. Tento předpoklad se potvrdil a u všech metod došlo, více či méně, k zlepšení výsledků. Výsledky na testovacích datech pro různé metody po eliminaci signálů ze skupiny s ostatními arytmiemi jsou uvedeny v souhrnné tabulce 9.9.

Tab. 9.9: Výsledky pro redukováná testovací data.

VÝSLEDKY METOD PO REDUKCI DAT				
Metoda	SE [%]	SP [%]	ACC [%]	PPV [%]
Prahovací metoda	82	73	78	76
SVM	86	80	83	81
KNN	96	70	83	77

Uvedené výsledky odpovídají prahovacím metodám, přičemž prahy pro jednotlivé příznaky nebyly nijak upraveny, dále klasifikátoru SVM s polynomiálním jádrem druhého stupně a klasifikátoru KNN s $k = 11$.

9.5 Kombinace metod

Na závěr byla vyzkoušena kombinace metod s rozhodovacím pravidlem. Pokud se alespoň dvě metody shodly na tom, že daný signál obsahuje fibrilace, byl takto označen.

Výsledky se markantně nezměnily, ale došlo k nárůstu u všech parametrů, byť jen o pár procent. Pro všechna testovací data byla senzitivita 80 %, specificita 71 %, přesnost 74 % a pozitivní prediktivní hodnota 60 %. Pro redukováná data byla senzitivita 94 %, specificita 72 %, přesnost 83 % a pozitivní prediktivní hodnota 77 %, výsledky byly velice blízké výsledkům metody KNN pro redukováná data.

9.6 Diskuze výsledků

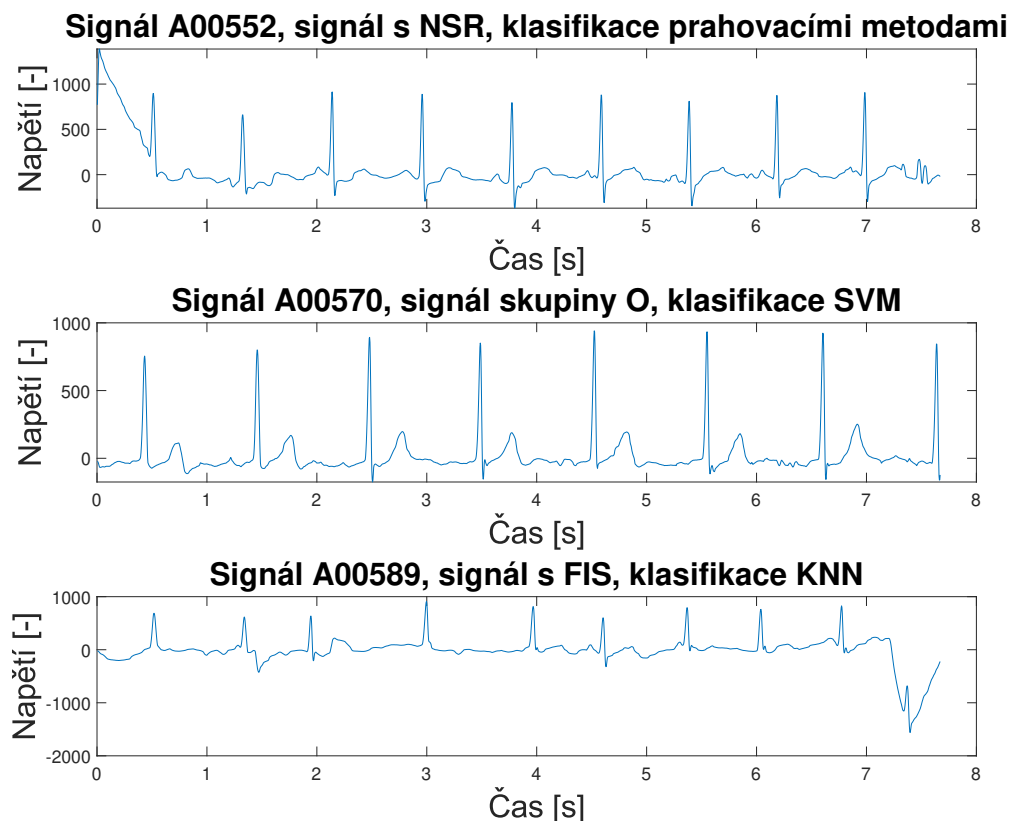
Zvolenými prahovacími metodami pro detekci fibrilací síní v signálech bylo dosaženo senzitivity 84 %, specificity 66 %, přesnosti 73 % a pozitivní prediktivní hodnoty 57 % pro trénovací data s osmi příznaky, CV, TPR, pNN20, pNN50, MAD, SE a nRMSSD a poměrem SD1/SD2. Po vybrání pěti příznaků, konkrétně pNN20, pNN50, MAD, SE, nRMSSD a optimalizování prahů pro tuto kombinaci se výsledky zlepšily, bylo dosaženo senzitivity 92 %, specificity 89 %, přesnosti 90 % a pozitivní prediktivní hodnoty 81 %.

Pro testovací data došlo k poklesu všech hodnot z důvodu prahů, které byly příliš optimalizovány pro trénovací data, proto nebyly tak účinné pro testovací data. V tomto případě algoritmus se všemi příznaky dosáhl nejvyšší senzitivity 80 %, specificity 62 %, přesnosti 68 % a pozitivní prediktivní hodnoty 51 %, algoritmus s pěti příznaky dosáhl senzitivity 86 %, specificity 75 %, přesnosti 79 % a pozitivní prediktivní hodnoty 63 %. Mezi špatně zařazené signály patřily nejčastěji signály s ostatními arytmiemi či signály s fibrilacemi, které nevykazovaly nepravidelnosti RR intervalů, ale například absenci vlny P.

Další použitou metodou byla SVM klasifikace. Ta byla v porovnání s prahovacími metodami při použití všech příznaků úspěšnější v odlišování ostatních arytmií, vznikala menší počet falešně negativních případů. Byly vyzkoušeny různé typy jader a jako nejlepší se jevil klasifikátor s polynomiálním jádrem druhého stupně. Tento klasifikátor dosáhl na testovací množině senzitivity 78 %, specificity 75 %, přesnosti 74 % a pozitivní prediktivní hodnoty 61 %.

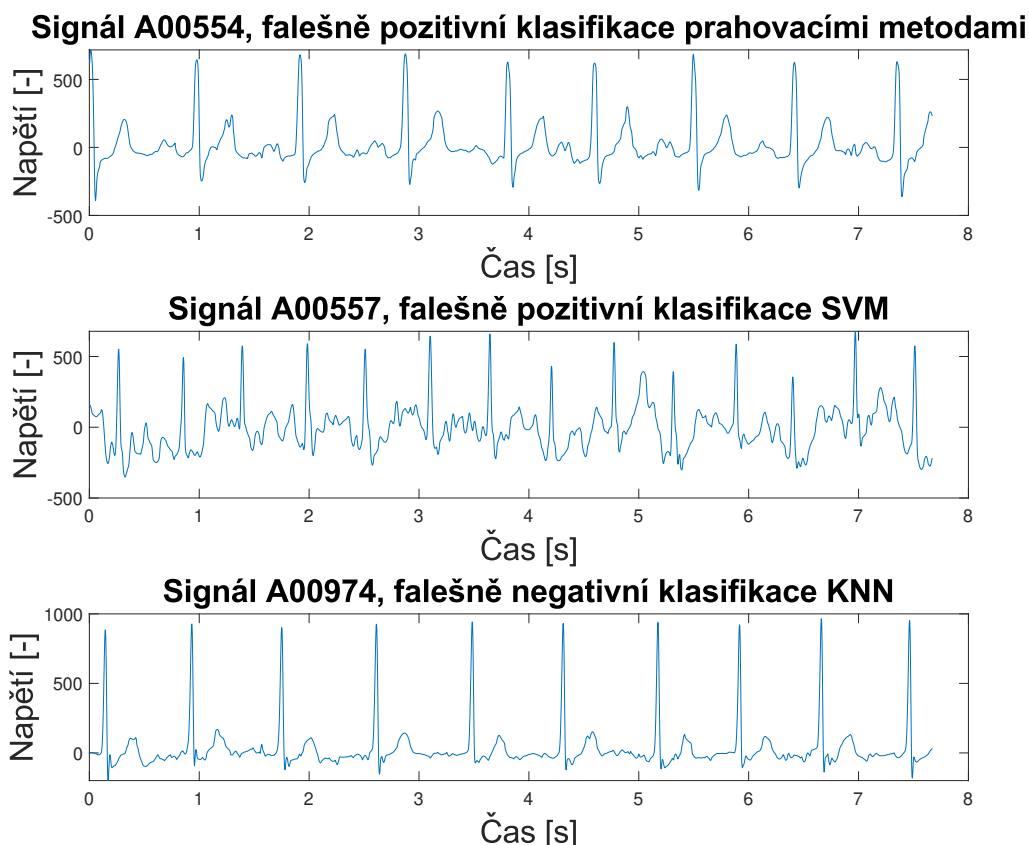
Poslední použitou metodou byla klasifikace KNN, která byla optimalizována na trénovací množině pomocí trojnásobné křížové validace, čímž byla ověřena generalizační schopnost natrénovaného klasifikátoru. Jako nejlepší ze všech vyzkoušených klasifikátorů s různým počtem sousedů se jevil klasifikátor s jedenácti sousedy, který dosahoval senzitivity 78 %, 72 %, 74 % a 58 %. Výsledky byly blízké SVM klasifikaci a lepší než výsledky prahovacích metod při všech příznacích.

Na obrázku 9.2 jsou zobrazeny tři grafy správně klasifikovaných signálů každou metodou. První graf zobrazuje signál s normálním sinusovým rytmem, který je správně klasifikován pomocí prahovacích metod. Na druhém grafu je vidět signál s přítomností jiné arytmie, který byl správně zařazen metodou SVM. Na posledním grafu je vidět patologický signál s fibrilacemi, který byl správně zařazen pomocí metody KNN. Na tomto signálu lze pozorovat hlavní příznaky fibrilací, tedy absence vlny P, silná nepravidelnost RR intervalů a f vlnky v některých úsecích.



Obr. 9.2: Ukázky správně klasifikovaných signálů.

Na obrázku 9.3 jsou zobrazeny tři grafy nesprávně klasifikovaných signálů každou metodou. První graf zobrazuje falešně pozitivní klasifikaci pomocí prahovacích metod. Signál sice není úplně fyziologický, nicméně nebyl anotován jako signál s fibrilacemi, nýbrž jako signál s jinou arytmií. Vizuálně vykazuje mírnou nepravidelnost RR intervalů, na základě této nepravidelnosti byl signál vyhodnocen jako pozitivní. P vlna je zachována, v tomto případě by tedy bylo vhodnější doplnit sadu příznaků o parametr, který by značil přítomnost či absenci P vlny. Druhý graf zobrazuje signál, jež se opět vyznačuje nepravidelností RR intervalů, signál je silně patologický a obsahuje mnoho drobných vlnků, které mohou připomínat f vlnky typické pro fibrilace, zároveň je těžko rozeznatelná P vlna, nicméně signál byl anotován jako signál s jinou arytmií, z tohoto důvodu je označen jako falešně pozitivní. Na posledním grafu lze vidět signál, který je označen jako negativní i přes to, že se v něm vyskytují příznaky fibrilací, jako například f vlnky či absence vlny P. Opět by v tomto případě mohl rozhodnout parametr přítomnosti či absence vlny P, neboť nepravidelnost RR intervalů zde není vizuálně příliš viditelná.



Obr. 9.3: Ukázky nesprávně klasifikovaných signálů.

Pro zajímavost byly na závěr z databáze vyřazeny signály s ostatními arytmiemi. Cílem bylo odpovědět na otázku, jak se změní výsledky po eliminaci těchto signálů. Ne vždy se totiž podařilo ostatní arytmie správně odlišit a vznikalo poměrně mnoho falešně negativních právě kvůli ostatním patologickým signálům, které výsledky zkreslovaly.

Byla také vyzkoušena kombinace těchto tří metod s rozhodovacím pravidlem. Došlo k drobnému zlepšení výsledků, nicméně vzhledem k tomu, že metody fungovaly se stejnými příznaky, povětšinou se na klasifikaci signálu shodly stejně dobře, případně stejně špatně, jejich kombinace tedy nepřinesla žádný velký úspěch.

Všechny výsledky se zlepšily ve všech parametrech. Po redukci dat dopadla nejlépe SVM klasifikace s polynomiálním jádrem druhého stupně, touto metodou bylo dosaženo senzitivity 86 %, specifity 80 %, přesnosti 83 % a pozitivní prediktivní hodnoty 81 %. Metoda KNN dosáhla největší senzitivity ze všech tří použitých algoritmů, konkrétně 96 %, ale vznikalo mnoho falešně negativních případů, kvůli kterým byla specifita pouze 70 %.

Největším kamenem úrazu těchto algoritmů jsou příznaky, které vycházejí z nepravidelnosti RR intervalů. Obecně by metody vylepšilo zahrnutí přítomnosti či absence vlny P do příznaků pro následné prahování či klasifikaci. Použitelné by byly i příznaky vycházející ze stacionární vlnkové transformace, které byly použity ve studii [7].

Po vyřazení ostatních arytmií sice došlo ke zvýšení parametrů, nicméně i tak byly v databázi signály s fibrilacemi, které nevykazovaly nepravidelnosti RR intervalů, z toho důvodu byly vyhodnoceny jako falešně negativní.

9.7 Porovnání výsledků s publikovanými studiemi

Porovnání výsledků této bakalářské práce se studiemi zahraničních autorů není relevantní, protože signály použité pro tuto konkrétní práci byly přímo poskytnuty z databáze UBMI a nebyla nalezena žádná volně přístupná studie používající tyto metody na těchto konkrétních datech, nicméně v literatuře bylo použití SVM klasifikace poměrně časté, stejně tak použití prahovacích metod [5, 6, 11, 24, 25, 26, 34].

Závěr

Cílem této bakalářské práce bylo provést literární rešerši v oblasti metod detekce FIS v signálech EKG a následné implementaci vybraných metod a vyhodnocení funkčnosti algoritmu.

V první kapitole jsou uvedeny základní informace o vzniku a charakteru síňových fibrilací. V druhé kapitole jsou uvedeny jednotlivé metody, které jsou v praxi využívány pro detekci FIS, od těch nejjednodušších po úvod do nejsložitějších metod využívajících metod hlubokého učení, neuronových sítí.

V praktické části bylo uvedeno předzpracování dostupných dat. Data pocházela z databáze UBMI, byla pořízena zařízením AliveCor s vzorkovací frekvencí 300 Hz a průměrná délka byla 35 vteřin. Jednalo se o celkem 300 signálů, z nichž 100 obsahovalo fibrilace síní, 100 bylo fyziologických s normálním sinusovým rytmem a 100 obsahovalo jinou arytmií. Celý tento dataset byl dodán ve dvou podsetech po 150 signálech, první dataset byl použit jako trénovací data a druhý jako testovací data, což zajišťovalo relevantnost výsledků, neboť metody byly optimalizovány na trénovacích datech a na datech testovacích pouze otestovány.

Prvním krokem po předzpracování dat byla detekce QRS komplexů pomocí hrubého odhadu kvadrátu obálky a následného prahování adaptivním prahem. K získání výsledků byl využit testovací software dostupný z UBMI. QRS komplexy byly detekovány se senzitivitou 97,56 % a pozitivní prediktivní hodnotou 99,58 %.

Po získání pozic QRS komplexů následovala samotná detekce FIS. K detekci byly použity tři metody a na závěr pak kombinace těchto metod s rozhodovacím pravidlem.

Pro použití vybraných metod byl rozhodující výběr příznaků. Původně deset příznaků vycházejících z nepravidelnosti RR intervalů bylo podrobeno statistické analýze v softwaru STATISTICA verze 13.5. Zde bylo na začátku zjištěno, které příznaky pro jednotlivé skupiny vykazují normální rozdělení, poté byly zvoleny statistické testy, které ověřovaly, zda jsou příznaky významné pro odlišení jednotlivých klasifikačních skupin. Pomocí statistických testů bylo původních deset příznaků zredukováno na výsledných osm. Tyto příznaky byly dále využity pro klasifikaci dat.

Jako první byla využita kombinace prahovacích metod. Jednotlivé příznaky byly prahovány nejvhodnějším prahem a na základě rozhodovacího pravidla byly jednotlivé signály klasifikovány jako signály s fibrilacemi nebo bez, případně s ostatními arytmiemi, tyto dvě skupiny byly v hodnocení sloučeny dohromady. Touto kombinací bylo dosaženo senzitivity 84 %, specifity 66 %, přesnosti 73 % a pozitivní prediktivní hodnoty 57 % na trénovací množině. Byl očekáván pokles hodnot těchto parametrů, především kvůli optimalizaci prahu na trénovací množině dat. Na testovacích datech dosáhl tento algoritmus senzitivity 80 %, specifity 62 %, přesnosti

68 % a pozitivní prediktivní hodnoty 51 %. Možným vylepšením této metody by bylo zabudování adaptivního prahu, který by se měnil dle hodnot parametrů. Po vybrání pěti příznaků a upravení prahů pro jejich kombinaci bylo dosaženo lepších výsledků, všechny parametry se zvýšily o několik procent.

Druhou použitou metodou byla klasifikace pomocí SVM klasifikátoru. Ten byl vytvořen pomocí již implementovaných funkcí v prostředí MATLAB R2020b. Po natrénování byl klasifikátor využit pro validaci na testovacích datech, přičemž bylo vyzkoušeno několik typů jader funkce. Data nejlépe klasifikoval klasifikátor s polynomiálním jádrem druhého stupně, který dosahoval senzitivity 78 %, specificity 75 %, přesnosti 76 % a pozitivní prediktivní hodnoty 61 %.

Poslední použitou metodou byla klasifikace pomocí klasifikátoru KNN, který byl vytvořen pomocí implementované funkce v prostředí MATLAB R2020b. Tato metoda byla optimalizována na trénovacích datech pomocí trojnásobné křížové validace. Data nejlépe klasifikoval klasifikátor s jedenácti sousedy, který dosahoval senzitivity 78 %, specificity 72 %, přesnosti 74 % a pozitivní prediktivní hodnoty 58 %.

Kombinace těchto tří metod nepřinesla žádnou významnou změnu ve výsledcích, především proto, že všechny metody data klasifikovaly na základě stejných příznaků a ve většině případů na klasifikaci shodly, nicméně bylo použito rozhodovací pravidlo, které daný signál označilo za patologický s fibrilacemi, pokud se na tom shodly alespoň dvě metody ze tří.

Pro zajímavost byla data z trénovací i testovací databáze zredukována a byly odstraněny signály s ostatními arytmiemi, aby bylo možno zjistit, jak moc jsou výsledky zkreslené těmito ostatními arytmiemi. Po této redukci byla nejúspěšnější metoda SVM, která dosahovala hodnot senzitivity 86 %, specificity 80 %, přesnosti 83 % a pozitivní prediktivní hodnoty 81 %.

Pro zlepšení výsledků by mohlo pomoci rozšířit příznaky o příznak značící přítomnost či absenci P vlny, jejíž absence je typická pro fibrilace síní. Některé signály s fibrilacemi totiž nevykazují velké rozdíly mezi jednotlivými RR intervaly, proto jsou nesprávně zařazeny mezi fyziologické signály. Dále by bylo možno vylepšit prahovací metody adaptivními prahy pro jednotlivé příznaky.

Literatura

- [1] LUKL, Jan. *Fibrilace síní*. Praha: Grada, 2009. ISBN 978-80-247-2768-4.
- [2] BENNETT, David H. *Srdeční arytmie: praktické poznámky k interpretaci a léčbě*. Praha: Grada, 2014. ISBN 978-80-247-5134-4.
- [3] YAGHOUBY, Farid et al. Towards automatic detection of atrial fibrillation: A hybrid computational approach. *Computers in biology and medicine* [online]. 2010, (40), 919-930 [cit. 2020-11-23]. Dostupné z: <https://www.sciencedirect.com/science/article/abs/pii/S0010482510001435?via%3Dihub>
- [4] KRUGER, G.H. et al. Bimodal classification algorithm for atrial fibrillation detection from m-health ECG recordings. *Computers in biology and medicine* [online]. 2019, (104), 310-318 [cit. 2020-11-23]. Dostupné z: <https://www.sciencedirect.com/science/article/abs/pii/S0010482518303780?via%3Dihub>
- [5] KENNEDY BSC, Alan at al. Automated detection of atrial fibrillation using R-R intervals and multivariate-based classification. *Journal of Electrocardiology* [online]. 2016 [cit. 2020-04-21]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S002207361630111X>
- [6] GHODRATI, Alireza, Bill MURRAY a Stephen MARINELLO. RR interval analysis for detection of Atrial Fibrillation in ECG monitors. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* [online]. 2008, **2008**, 601-604 [cit. 2020-11-26]. Dostupné z: <https://ieeexplore.ieee.org/document/4649224>
- [7] MARŠÁNOVÁ, Lucie, Andrea NĚMCOVÁ a Radovan SMÍŠEK AT AL. Advanced P Wave Detection in Ecg Signals During Pathology: Evaluation in Different Arrhythmia Contexts. *Scientific Reports* [online]. 2019, **2019**(9) [cit. 2020-12-22]. Dostupné z: <https://doi.org/10.1038/s41598-019-55323-3>
- [8] TATENO, K. a L. GLASS. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ARR intervals. *Med. Biol. Eng. Comput.* [online]. **2001**(39), 664-671 [cit. 2020-12-22]. Dostupné z: http://www.medicine.mcgill.ca/physio/glasslab/pub_pdf/automatic_2001.pdf

- [9] NIST/SEMATECH e-Handbook of Statistical Methods. *NIST/SEMATECH e-Handbook of Statistical Methods* [online]. 2012 [cit. 2020-12-22]. Dostupné z: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm>
- [10] TANG, SC., PW. HUANG a CS. HUNG. Identification of Atrial Fibrillation by Quantitative Analyses of Fingertip Photoplethysmogram. *Scientific Reports* [online]. 2017, **2017**(7) [cit. 2020-12-22]. Dostupné z: <https://www.nature.com/articles/srep45644>
- [11] KRIVOSHEI, Lian et al. Smart detection of atrial fibrillation. *EP Europace* [online]. 2017, 5 May 2017, , 753-757 [cit. 2020-04-21]. Dostupné z: <https://academic.oup.com/europace/article/19/5/753/2952343>
- [12] DUGAR, Diva. Skew and Kurtosis: 2 Important Statistics terms you need to know in Data Science. In: *Codeburst* [online]. [cit. 2020-12-22]. Dostupné z: <https://tinyurl.com/yyxcbt5v>
- [13] DASH, S. et al. Automatic Real Time Detection of Atrial Fibrillation. *Annals of Biomedical Engineering* [online]. **2009**(9), 1701-1709 [cit. 2020-12-22]. Dostupné z: <https://pubmed.ncbi.nlm.nih.gov/19533358/>
- [14] GOLÍŇSKA, Agnieszka Kitlas. Poincare Plots in Analysis of Selected Bio-medical Signals. *Sciend* [online]. 2013, **35**(48) [cit. 2020-12-22]. ISSN 2199-6059. Dostupné z: <https://content.sciendo.com/view/journals/slgr/35/1/article-p117.xml?language=en>
- [15] SEVŠEK, France. Shape Analysis of Heart Rate Lorenz Plots. *Cellular and Molecular Biology Letters* [online]. 2002, **7**(1), 159-161 [cit. 2020-12-22]. Dostupné z: https://www.researchgate.net/publication/11422638_Heart_rate_variability--a_shape_analysis_of_Lorenz_plots
- [16] ANDERSSON, Oskar et al. A 290 mV Sub- VT ASIC for Real-Time Atrial Fibrillation Detection. *IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS* [online]. 2015, **9**(3), 377 - 386 [cit. 2020-12-22]. Dostupné z: <https://ieeexplore.ieee.org/document/6926866>
- [17] HUANG, Chao et al. A Novel Method for Detection of the Transition Between Atrial Fibrillation and Sinus Rhythm. *IEEE Transactions on Biomedical Engineering* [online]. 2011, **58**(4), 1113 - 1119 [cit. 2020-12-23]. Dostupné z: <https://ieeexplore.ieee.org/document/5657251>
- [18] BASTL, P., J. KUČERA a D. LÉWOVÁ. *Metody strojového učení* [online]. [cit. 2020-12-28]. Dostupné z: <http://tydenvedy.fjfi.cvut.cz/2003/cd/prispevky/sbpdf/strojuc.pdf>

- [19] FAUST, Oliver, Edward CIACCIO a Rajendra ACHARYA. A Review of Atrial Fibrillation Detection Methods as a Service. *Int J Environ Res Public Health* [online]. 2020, **17**(9) [cit. 2020-12-28]. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7246533/>
- [20] MUSIL, Tomáš. *Naivní Bayesův klasifikátor* [online]. [cit. 2020-12-28]. Dostupné z: <http://su.tomasm.cz/pdf/bayes.pdf>
- [21] MARTIS, Roshan Joy et al. Automated detection of atrial fibrillation using Bayesian paradigm. *Knowledge-Based Systems* [online]. 2013, **54** [cit. 2020-12-28]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0950705113002943>
- [22] ZHÁNĚL, DR., Doc. RNDr. Jiří a kolektiv. *Aplikace výzkumných metod v kinantropologii* [online]. Brno, 2014 [cit. 2020-12-28]. ISBN 978-80-210-6793-6. Dostupné z: <https://www.fsps.muni.cz/emuni/data/reader/book-34/05.html>
- [23] HOUDEK, Michal, Tomáš SVOBODA a Tomáš PROCHÁZKA. *Klasifikace podle nejbližších sousedů Nearest Neighbour Classification [k-NN]* [online]. 2001 [cit. 2020-12-28]. Dostupné z: http://cmp.felk.cvut.cz/cmp/courses/recognition/zapis_prednasky/zapis_01/4/rpz4.pdf
- [24] ASGRARI, S., A. MEHRNIA a M. MOUSSAVI. Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Computers in Biology and Medicine Volume* [online]. 2015, **60**, 132-142 [cit. 2020-12-28]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0010482515000839>
- [25] MOHEBBI, M. a H. GHASSEMIAN. Detection of atrial fibrillation episodes using SVM. *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* [online]. 2008, , 177-180 [cit. 2020-12-28]. Dostupné z: <https://ieeexplore.ieee.org/document/4649119>
- [26] APARNA, P. a K. M. SHARMA. Detection of A Fib and its Classification using SVM. *2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* [online]. 2020, , 116-120 [cit. 2020-12-28]. Dostupné z: <https://ieeexplore.ieee.org/document/9074888>
- [27] CARVAJAL, Diana a Peter ROWE. Sensitivity, specificity, predictive values, and likelihood ratios. *Pediatrics in review / American Academy of Pediatrics* [online]. 2010, **31**(12), 511-513 [cit. 2020-12-31]. Dostupné

- z: https://www.researchgate.net/publication/49650721_Sensitivity_specificity_predictive_values_and_likelihood_ratios
- [28] BARATLOO, A. et al. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emerg (Tehran)* [online]. 2015, , 48-49 [cit. 2020-12-31]. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4614595/>
 - [29] Neuronové sítě. *Mendelova univerzita v Brně* [online]. [cit. 2020-12-31]. Dostupné z: https://is.mendelu.cz/eknihovna/opory/zobraz_cast.pl?cast=21471
 - [30] VÍTEK, M. a J. KOZUMPLÍK. CSE QRS TESTER; Software pro testování detektorů QRS na databázi CSE. *Ústav biomedicínského inženýrství Vysoké učení technické v Brně* URL: <http://www.ubmi.feec.vutbr.cz/vyzkum-a-vyvoj/produkty>. (software)
 - [31] FRIEDRICH, Václav, Pavel HRADECKÝ, Šárka MICHALCOVÁ a Marek POMP. *Vybrané statistické metody*. Ostrava: VŠB-TU Ostrava, 2018. ISBN 978-80-248-4108-3.
 - [32] *Neparametrické testy* [online]. [cit. 2021-5-27]. Dostupné z: <https://cit.vfu.cz/statpotr/POTR/Teorie/Predn4/MannWhit.htm>
 - [33] *Parametrické testy*. [online]. [cit. 2021-5-27]. Dostupné z: <https://cit.vfu.cz/stat/FVHE/Teorie/Predn3/ttest.htm>
 - [34] RESIANDI, K., ADIWIJAYA a D. Q. UTAMA. Detection of Atrial Fibrillation Disease Based on Electrocardiogram Signal Classification Using RR Interval and K-Nearest Neighbor. *2018 6th International Conference on Information and Communication Technology (ICoICT)* [online]. , 501-506 [cit. 2021-5-27]. Dostupné z: doi:10.1109/ICoICT.2018.8528737.
 - [35] Support vector machines (SVM). *Masarykova univerzita v Brně* [online]. [cit. 2021-5-27]. Dostupné z: https://is.muni.cz/el/1433/podzim2006/PA034/09_SVM.pdf

A Popis postupu ověření navržených algoritmů

A.1 Navržené funkce

Bylo naprogramováno pět funkcí *detektor R*, *priznaky*, *prahovani*, *KNN*, *SVM* a dvě funkce *Main*, *GetScoreFIS*, byly dodány z UBMI.

Funkce *detektor R* slouží pro nalezení QRS komplexů. Funkce *priznaky* slouží pro extrakci příznaků pro metody KNN a SVM. Zbylé naprogramované tři funkce slouží pro prahování, metodu KNN a metodu SVM. Funkce dodané UBMI slouží pro ověření funkčnosti prahovacích metod.

Pro ověření bylo vybráno 10 trénovacích a 10 testovacích signálů, na kterých je možno ověřit funkčnost navržených algoritmů. Výstupem všech funkcí jsou hodnoty SE, SP, ACC a PPV.

A.2 Ověření prahovacích metod

Pro spuštění slouží funkce *Main*, která byla včetně funkce *GetScoreFIS* dodána UBMI, pouze dle potřeb upravena, pro ověření funkčnosti stačí zadat příkaz ve tvaru:

```
[ se, sp, acc, ppv ] = Main( cesta, reference )
```

kde *filePath* je cesta k datům (trénovacím nebo testovacím) a *reference* je soubor ve formátu *.csv s referencemi pro daná data. Konkrétně se dají prahovací metody ověřit pomocí příkazů:

```
[ se, sp, acc, ppv ] = Main( 'dataTrain', 'dataTrain.csv' )  
[ se, sp, acc, ppv ] = Main( 'dataTest', 'dataTest.csv' )
```

A.3 Ověření metody SVM

Ověření metody SVM probíhá příkazem v obecném tvaru:

```
[ SE, SP, ACC, PPV ] = SVM( cestaTrain, cestaTest, referenceTrain,  
referenceTest )
```

kde *cestaTrain* a *cestaTest* jsou cesty k datům a *referenceTrain* a *referenceTest* jsou reference pro příslušná data. Konkrétně lze metodu KNN ověřit příkazem:

```
[ SE, SP, ACC, PPV ] = SVM( 'dataTrain', 'dataTest', 'dataTrain.csv',  
'dataTest.csv' )
```

A.4 Ověření metody KNN

Pro ověření metody KNN stačí zadat příkaz v obecném tvaru:

```
[ SE, SP, ACC, PPV ] = kNN( cestaTrain, cestaTest, referenceTrain,  
referenceTest )
```

kde *cestaTrain* a *cestaTest* jsou cesty k datům a *referenceTrain* a *referenceTest* jsou reference pro příslušná data. Konkrétně lze metodu KNN ověřit příkazem:

```
[ SE, SP, ACC, PPV ] = kNN( 'dataTrain', 'dataTest', 'dataTrain.csv',  
'dataTest.csv' )
```

Na řádku č. 58 je možné měnit počet sousedů, pro ověřovací případ je zde nastaveno $k = 1$. nikoli $k = 11$, jelikož je k dispozici omezený počet dat.

B Obsah adresáře přiloženého v IS VUT

V souboru přiloženém v informačním systému VUT se nachází 7 funkcí, soubor readme v poznámkovém bloku a dvě složky dataTrain a dataTest, které slouží pro trénování a testování navržených metod.

```
Behuncikova
├── kNN.m
├── GetScoreFIS.m
├── SVM.m
├── detektor_R.m
├── readme.txt
├── Main.m
├── prahovani.m
├── priznaky.m
├── dataTrain
│   ├── A00161.mat
│   ├── dataTrain.csv
│   ├── A00509.mat
│   ├── A00520.mat
│   ├── A00081.mat
│   ├── A00079.mat
│   ├── A00519.mat
│   ├── A00542.mat
│   ├── A00159.mat
│   ├── A00080.mat
│   └── A00158.mat
└── dataTest
    ├── A01024.mat
    ├── A00693.mat
    ├── A01013.mat
    ├── A01005.mat
    ├── A00633.mat
    ├── A01019.mat
    ├── A00631.mat
    ├── A00696.mat
    ├── A00697.mat
    ├── dataTest.csv
    └── A00632.mat
```